Diogo Filipe Neves de Oliveira

Master of Science

Remote Sensing and Data Fusion for Eucalyptus Trees Identification

Dissertation submitted in partial fulfillment
of the requirements for the degree of

Master of Science in
Computer Science and Engineering

Adviser: Maria Rita Sarmento de Almeida Ribeiro,
Invited Associate Professor,
Electrical Engineering Department (DEEC), FCT-UNL

Co-adviser: Carlos Augusto Isaac Piló Viegas Damásio,
Associate Professor,
Department of Computer Science (DI), FCT-UNL

Examination Committee

Chair:
Rapporteurs:
Members:

FACULDADE DE
CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

September, 2019

Remote Sensing and Data Fusion for Eucalyptus Trees Identification

This document was created using the (pdf)LaTeX processor, based on the "novathesis" template[1], developed at the Dep. Informática of FCT-NOVA [2].
[1] https://github.com/joaomlourenco/novathesis [2] http://www.di.fct.unl.pt

À minha família e amigos

Acknowledgements

Gostaria de agradecer a todos aqueles que contribuíram e ajudaram de forma decisiva para realização desta dissertação, que marca o fim de uma etapa muito importante da minha vida.

Primeiro, quero agradecer à minha orientadora, Professora Rita Ribeiro por todos os conhecimentos que me transmitiu, pela paciência e dedicação que demonstrou ao longo de todo o desenvolvimento. Quero agradecer também ao meu co-orientador, professor Carlos Damásio, pela orientação e conselhos prestados.

A todo o grupo do CA3-UNINOVA, nomeadamente os professores José Fonseca e André Mora, por me terem auxiliado sempre que necessário e me fazerem sempre sentir parte do grupo, ao longo de todo o período.

Um agradecimento profundo à minha família pelo apoio incondicional ao longo destes anos. À minha mãe Elisabete Neves, a minha maior inspiração, ao meu pai Henrique Oliveira, ao meu irmão Ricardo, por ter estado sempre disponível para tudo o que precisasse e por toda a ajuda e conselhos que me tem dado ao longo da minha vida. A todos os meus tios e primos, especialmente à minha tia Sónia, que me acolheu e ajudou em tudo durante estes 5 anos. Por último e não menos importante, uma palavra de apreço e gratidão a todos os meus avós que me ajudaram de forma incalculável ao longo da minha vida, sendo determinantes para a conclusão desta importante etapa.

A todos os colegas de curso que conheci ao longo do meu percurso, por tudo o que passámos e pela amizade criada, que me ajudou a ultrapassar os piores momentos destes 5 anos. De todos os colegas que conheci e amigos com quem passei estes anos, durante o meu percurso académico, quero destacar os meus grandes amigos: André Victorino, André Pereira, Diogo Fernandes, Manuel Barbas, Paulo Dias, Pedro Cabaço e Tito Ferreira. A todos os meus amigos de sempre, que sempre me ajudaram em qualquer situação, uma grande palavra de gratidão. Em cada um de vocês levo amizade para a vida.

Por último quero agradecer à Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa pela forma como me acolheu nestes últimos 5 anos e pela disponibilidade apresentada para o que precisasse.

Abstract

Satellite remote sensing is supported by the extraction of data/information from satellite images or aircraft, through multispectral images, that allows their remote analysis and classification. Analyzing those images with data fusion tools and techniques, seem a suitable approach for the identification and classification of land cover.

This land cover classification is possible because the fusion/merging techniques can aggregate various sources of heterogeneous information to generate value-added products that facilitate features classification and analysis. This work proposes to apply a data fusion algorithm, denoted FIF (Fuzzy Information Fusion), which combines computational intelligence techniques with multicriteria concepts and techniques to automatically distinguish Eucalyptus trees, in satellite images To assess the proposed approach, a Portuguese region, which includes planted Eucalyptus, will be used. This region is chosen because it includes a significant number of eucalyptus, and, currently, it is hard to automatically distinguish them from other types of trees (through satellite images), which turns this study into an interesting experiment of using data fusion techniques to differentiate types of trees.

Further, the proposed approach is tested and validated with several fusion/aggregation operators to verify its versatility. Overall, the results of the study demonstrate the potential of this approach for automatic classification of land types.

Keywords: Satellite images, data fusion, remote sensing, classification, fuzzy logic, computational intelligence, eucalyptus classification

Resumo

A deteção remota de imagens de satélite é baseada na extração de dados / informações de imagens de satélite ou aeronaves, através de imagens multiespectrais, que permitem a sua análise e classificação. Quando estas imagens são analisadas com ferramentas e técnicas de fusão de dados, torna-se num método muito útil para a identificação e classificação de diferentes tipos de ocupação de solo.

Esta classificação é possível porque as técnicas de fusão podem processar várias fontes de informações heterogéneas, procedendo depois à sua agregação, para gerar produtos de valor agregado que facilitam a classificação e análise de diferentes entidades - neste caso a deteção de eucaliptos. Esta dissertação propõe a utilização de um algoritmo, denominado FIF (Fuzzy Information Fusion), que combina técnicas de inteligência computacional com conceitos e técnicas multicritério. Para avaliar o trabalho proposto, será utilizada uma região portuguesa, que inclui uma vasta área de eucaliptos. Esta região foi escolhida porque inclui um número significativo de eucaliptos e, atualmente, é difícil diferenciá-los automaticamente de outros tipos de árvores (através de imagens de satélite), o que torna este estudo numa experiência interessante relativamente ao uso de técnicas de fusão de dados para diferenciar tipos de árvores.

Além disso, o trabalho desenvolvido será testado com vários operadores de fusão/agregação para verificar sua versatilidade. No geral, os resultados do estudo demonstram o potencial desta abordagem para a classificação automática de diversos tipos de ocupação de solo (e.g. água, árvores, estradas etc).

Palavras-chave: Imagens de satélite, fusão de dados, deteção remota, classificação, lógica difusa, inteligência computacional, classificação de eucaliptos

Contents

List of Figures

List of Tables

Acronyms

AI	Artificial Inteligence
ANN	Artificial Neural Network
CA3	Computational Intelligence Research Group
CART	Classification and Regression Tree
CI	Computational Intelligence
COS	Soil Occupancy Chart
DEM	Digital Elevation Model
DGT	General Directory of Territory
DT	Decision Tree
EPSG	European Petroleum Survey Group
ESA	European Space Agency
FIF	Fuzzy Information Fusion algorithm
FIS	Fuzzy Inference System
GIS	Geographic Information System
GML	Gaussian Maximum Likelihood
GPS	Global Positioning System
ICNF	Institute for Conservation of Nature and Forestry
IFOV	Instantaneous Field of View
INE	National Institute of Statistics
IPSTERS	IPSentinel Terrestrial Enhanced Recognition System
NUT	Nomenclature of Territorial Units for Statistical Purposes

SVM Support Vector Machine

tiff Tagged Image File Format

UAV Unmanned Aerial Vehicle *

UTM Universal Transverse Mercator

WGS World Geodetic System

Chapter 1

Introduction

The need for identification/classification of land cover (e.g. forests, water bodies, etc.) has become a very important subject over the years to allow public entities to monitor and manage forests, water bodies and so forth (Desclee et al. 2006). Existing tools and techniques are not enough since it is a hard task to analyze and distinguish land types, particularly different types of forests, because they can be very similar on satellite images due to their resolution. One Portuguese National entity working on identification and classification of the Portuguese territory is called DGT- General Directory of Territory http://www.dgterritorio.pt/. DGT provides the Portuguese Continental cartography and using COS (Soil Occupation Charts), it divides the maps into polygons of different types of land cover.

Although this service is very useful, it has two main problems. First, it is not an automatic process, hence it takes a long time to execute (so far only the 2015 COS is available). Second, due to satellite image resolution and diversity of land(e.g. forests, roads, houses) it is a rather difficult and imprecise process to identify/classify their classes/types. For example, it is easy to understand how difficult it is to distinguish, without doubt, the type of forest (e.g. eucalyptus trees), through images. Therefore, to automatically classify the classes/types of land is of paramount importance to improve monitoring of unlawful landscape changes or supporting precision agriculture (Taylor et al. 2000; Antrop 2004).

Land cover is the observed (bio) physical cover on the earth's surface, which include roads, grass, trees, water, soil, among others. It has an intrinsic characteristic, that is its natural indeterminacy, since land cover can have many different meanings. For example, in the United Kingdom areas without any trees can be sometimes classified as forest, while areas with slow-growing trees might not be classified as forest (Comber et al. 2005). There are two principal methods for capturing information about land cover: field survey and analysis of remotely sensed imagery (proposed work). Land change models can also

be built from those types of data to assess future shifts in land cover (Ross et al. 2006).

It is essential to implement systems that enable an automatic identification of satellite image characteristics, without human intervention, with greater certainty as possible to allow improvements on earth observation related topics (Emery et al. 2003; Davidson et al. 2007) . Satellite image identification goes through several processes , starting from the capture of data from satellites (digital images), to analyzing and identify the types of terrain (bodies of water, grass, fields of cultivation, forests, etc.) in both physical and chemical terms (Baumgardner et al. 1986; Ben-Dor et al.1997; Sumfleth et al. 2008; Ge et al. 2011).

Remote Sensing processes have been growing in the modern information society. This data is fundamental to monitor the Earth as an entire system, and several classification techniques can be applied to different areas. This type of image data can be processed by algorithms and other complex procedures, with the objective of making the best possible decision in terms of classification and monitoring of land territory (Gallego et al. 2012; Ribeiro et al. 2014; Lavreniuk et al. 2016;)

There are many types of terrains and many existing forest types, however, this work focus on eucalyptus trees because its plantation area is increasing in Portugal, in relation to other species, mainly because it is the main supplier of raw material for cellulose industries and provide a quick return on investment (Garcia 2017). According to the 6th National Forest Inventory, forests occupies near 35% of the total Portuguese soil (ICNF, 2017), which of those 26% are eucalyptus trees (INE 2017). This is a national challenge because planting eucalyptus is not a good choice on the long term, due to their high consumption of water and it interferes with the life of neighboring vegetation species (Liu et al. 2010).

For regulating the Portuguese Eucalyptus area, there is a Decree-Law No 96/2013, which constitutes the legal instrument to simplify and limit the plantation licensing of forest species and approves the legal regime for forestry management planning and intervention; however, it is considered insufficient to safeguard environmental issues (Garcia 2017). Since most remote sensing work on eucalyptus species identification is manual and there are few studies using some kind of data fusion processes (Ali et al. 2008; Haywood et al. 2011), this study can be a useful work base, and to contribute to future investigation.

Region of Study

The coordinate system used on the satellite images contains two main systems: The World Geodetic System (WGS), which is a standard in cartography geodesy, used in Sentinel 2 data; the Universal Transverse Mercator (UTM) system, which assigns coordinates to locations on the Earth's surface, and divides it into 60 zones. The definition of the coordinate systems is fundamental and the last revision WGS84 (1984) is the reference coordinate system used by the Global Positioning System (GPS). The acquired satellite image (Sentinel 2) are organized in ortho-rectified tiles of 100 km x 100 km in UTM WGS84 projections and has as coordinate reference system: *EPSG:32629 - WGS 84 / UTM zone*

29N. The European Petroleum Survey Group (EPSG) publishes a database of coordinate system information and Figure 1.1 depicts the chosen study region.

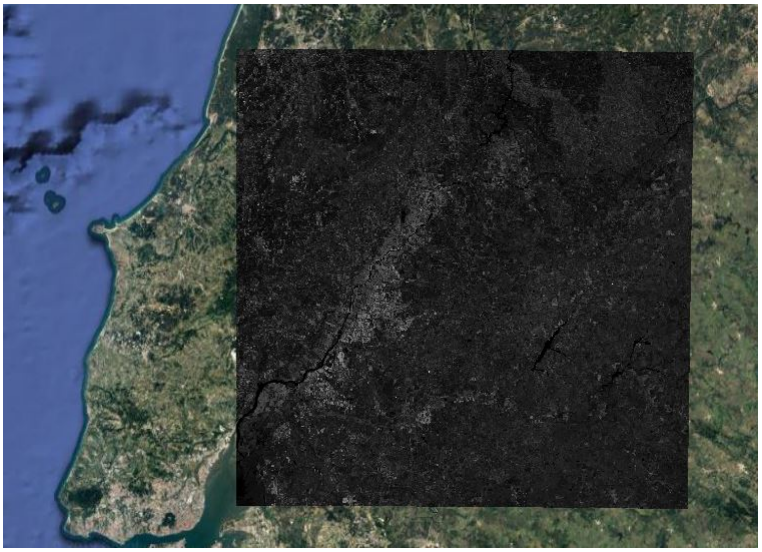


Figure 1.1: The study region highlighted in gray colors (Source: QGIS 3.4.5)

This was the chosen area because is an area of Portugal with a large amount of eucalyptus trees. That can be seen in Figure 1.2, where the red spots represents eucalyptus trees. This information is provided by DGT, through COS data.

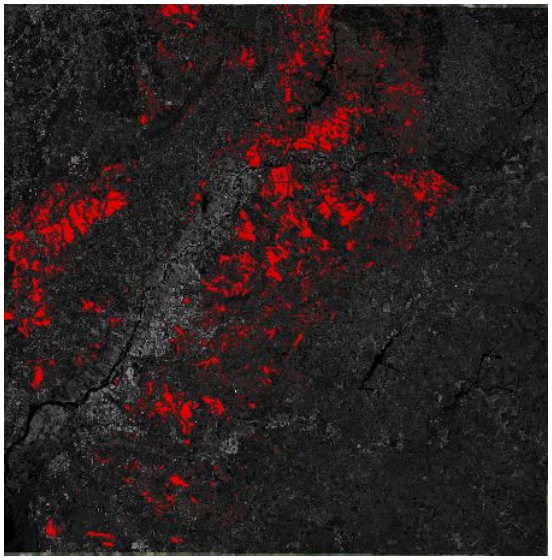


Figure 1.2: Amount of eucalyptus trees inside the study region, represented in red color (Source: QGIS 3.4.5)

To enable a more thorough study of the classification/identification of Eucalyptus and due to computational limitations the study region was further reduced as well as the selected areas for testing and validating the data fusion approach, which are detailed in Chapter 4.

1.1 Problem Description

Nowadays, land cover classification methods are really time consuming when done manually. The need to turn this process automatic is of great relevance for monitoring and management of land changes. Furthermore, as in any other scientific area, there is always a search for new methods or techniques, to improve the accuracy of results or facilitate information acquisition.

This work was performed under the IPSTERS project (IPSentinel Terrestrial Enhanced Recognition System), which main objective is exploring the application and limitations of AI (artificial intelligence) algorithms for processing large quantities of remotely sensed data, to produce, almost automatically, level-3 products for land applications such as precision agriculture, monitoring and management of forests etc. (http://www.ca3-uninova.org/project_ipsters). Within IPSTERS the work of this dissertation has as two objectives: first, to test a data fusion approach to automatically identify a specific type of tree (Eucalyptus tree); second to demonstrate the versatility and efficacy of a data fusion algorithm, denoted FIF (Ribeiro et al. 2014) for Sentinel 2 satellite remote sensing. Specifically, since COS sometimes does not discriminate (when it is smaller then 1 ha) between vegetation and, for instance roads, this work shows that with the AI approach it is possible to discriminate.

The aim is to support an intelligent classification of eucalyptus, by using computational intelligence techniques and data fusion processes for merging various sources of information, as described extensively in Chapter 5.

1.2 Main Contributions

The proposed solution main objective is to contribute to the field of remote sensing, more specifically to land cover classification, using a relatively recent data fusion algorithm (FIF) (Ribeiro et al.2014). We aim to demonstrate how suitable this approach is for land cover classification, specifically to classify Eucalyptus, and highlight the approach potential for classification of other land types.

1.3 Document Organization

Beyond this introductory chapter, the document is structured as follows:

Related Work

This chapter present the theoretical concepts regarding this subject and the current work being developed. We describe the most relevant aspects behind remote sensing, classification methodologies and technologies related to this work and its relevant literature.

Data Fusion Approach

In chapter 3 is described in detail the concepts behind the proposed approach.

System Development

In chapter 4 is described in detail the approach, how the created system was developed and how the data was prepared.

Case Study

In chapter 5 are described the several steps of the development of the performed work and are presented the obtained results, along with discussion of these results.

Conclusion and Further Work

Chapter 6 presents the conclusion for this work and further work that could be done as an extension of the developed approach. Also are described some points that can be improved.

CHAPTER 2

RELATED WORK

This chapter presents the main background concepts underlying the proposed approach: data fusion; remote sensing; fuzzy logic; membership functions; terrain feature classification.

2.1 Remote Sensing

Remote sensing systems importance derives from their ability to processing, monitoring, analysing, as well as predicting, the chemical, biological and physical aspects of the Earth (Wilson et al. 2013).

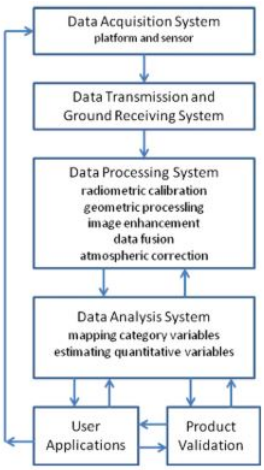


Figure 2.1: Components of remote sensing systems (Wilson et al. 2013).

When electromagnetic energy hits a component (soil, for example), there occurs an interaction that is called reflection, absorption and/or transmission of the radiation. This

is a crucial topic for remote sensing since these systems use the reflected information as captured data, i.e, when the sun light strikes a material, reflected data is the portion returned to the sensor system.

2.1.1 Spectral Information

The quantity of reflected information will directly influence the information captured by the sensing system, therefore, it will have impacts in many factors such as the nature of the observed material and where in the electromagnetic spectrum the measurement is being taken. Spectral information corresponds to the variation of reflectance regarding a material, in form of wavelengths, i.e., reflectance is a function of wavelength.

Figure 2.2 depicts Earth surface materials spectral reflectance signatures of the most common types: soil, green vegetation and water. The figure clearly shows how different their reflectance patterns are, over the different wavelengths.

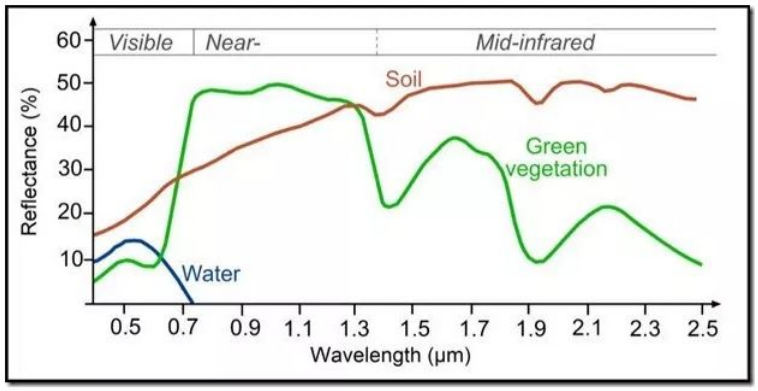


Figure 2.2: Reflectance of different surfaces. Source: (seos-project.eu)

As can be seen in Figure 2.2. , reflection levels at the visible region (0.4–0.7 μm) are relatively low and very close to each other, which generates spectral response patterns only in some specific cases and sometimes not enough (Eastman 2003). For example, in the case of vegetation, when the spectrum contains more information (after visible spectrum frontier), regarding infrared wave- lengths (>0.7μm), we can see that the signatures are more distinct in this region, which facilitates the identification and classification of different classes of vegetation.

The most used ranges for earth remote sensing are the visible/infrared range, which falls between about 0.4 and 12 μm, and the microwave range (between 30 and 300 mm). At microwave wavelengths, it is more common to use frequency and not wavelength to describe ranges of importance. So, the microwave range (30 to 300 mm) corresponds to frequencies between 1 GHz and 10 GHz. In the field of atmospheric remote sensing, common frequencies are in the range of 20 GHz to 60 GHZ (Richards et al. 1999).

The analysis and classification of spectral response signatures is the basis of land cover classification, regarding remote sensing systems (Huete 2004). Therefore, sensor systems, including multiple bands for capturing all the information are of paramount importance for capturing and analyzing reflectance patterns, not only in visible regions , but in other regions of the electromagnetic spectrum. Sentinel 2, for example, has thirteen bands, where only 4 are dedicated to the visible region of the spectrum. More on this subject will be discussed in section 2.1.2.2.

2.1.2 Remote Sensor Platforms

With the improvements made in technology over the years, a large variety of platforms arisen for the capture of remotely sensed data (Toth et al. 2016). It started with aerial photography, in the nineteenth century, to latest Unmanned Aerial Vehicle (UAV) remote sensing (Madry 2013).

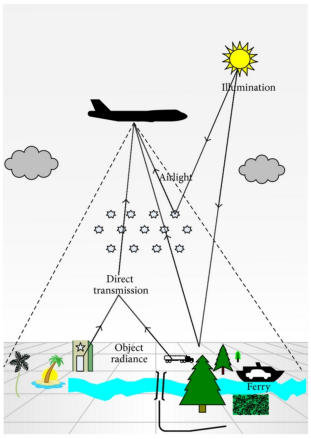


Figure 2.3: Model for airborne polarization remote sensing (Liang et al. 2016).

Nowadays, when we talk about remote sensing platforms, inevitably we mean satellite remote sensing, which started with Landsat-1 missions, back in 1972 (Simonetti et al. 2014). In1979, Seasat-1 become the first RADAR imaging satellite and started the new domain of remote sensing (Ouchi 2013).

Recently, satellite remote sensing platforms have witnessed many promising and exciting new developments. Examples of those are new higher spatial resolution optical and radar systems, hyperspectral sensors and important derived-products like digital elevation models (DEM). All of these improvements in remote sensing enabled the development of new processing techniques using machine learning techniques (Ali et al. 2015), which made satellite based systems very important for land cover classification (Pisani et al. 2014; Ahmed et al. 2015).

2.1.2.1 Satellite Based Systems

With more frequent usage of satellite systems for remote sensing, the need to broadcast imagery in digital form and the aim to produce/retrieve highly consistent imagery lead to the development of satellite-based scanning systems, as main format for the capture of remotely sensed data, and its constant improvements (Eastman 2003). In Figure 2.4 is depicted how a system like this works.

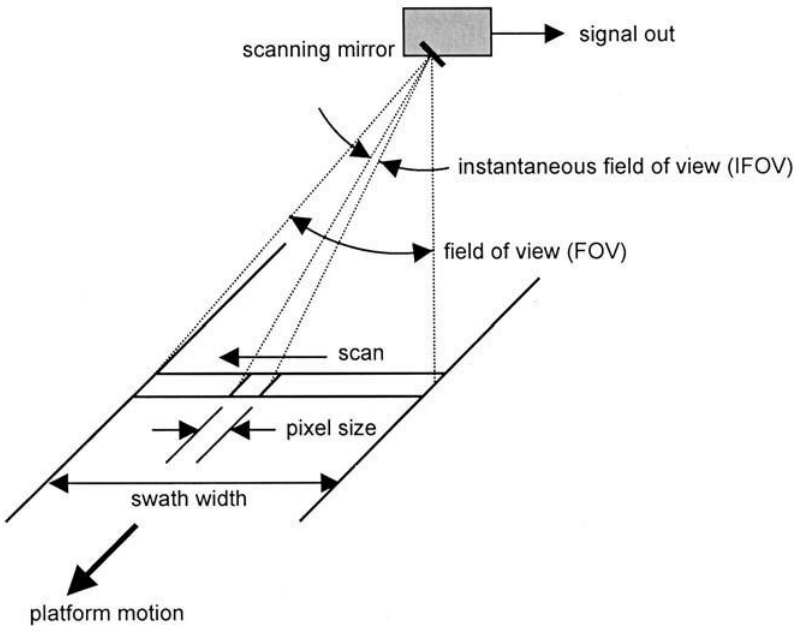


Figure 2.4: Mechanism of satellite-based systems (Daneshgar 2015).

The essential logic behind scanning sensors is the usage of a sweeping mechanism applied to a small field of view, known as instantaneous field of view (IFOV). This procedure is made in a west to east direction at the same time the satellite is moving in a north to south direction. More technically, a rotating mirror captures the energy that is allocated in the IFOV and then separates it into its spectral components.

After, there is an arrangement in the obtained spectrum path to allow electrical measurements of the amount of energy detected in various parts of the electromagnetic spectrum, provided by photoelectric detectors. As the scan moves (west to east), these photoelectric detectors are polled to gather a set of readings along the mentioned scan and. these procedures form the columns along one row of a set of raster images, one for each detector (Eastman 2003). In parallel, the satellite moves (north to south) and positions the system to detect the next row, and so forth. This process allows the production of a set of raster images as a record of reflectance over a range of spectral bands.

There are three important characteristics of satellite images: spatial resolution, spectral resolution and temporal resolution. Spatial resolution concerns the size of the ground area that is encapsulated by one data value in the set of images (IFOV). Spectral resolution

specifies the number and width of the spectral bands that the satellite sensor detects. Temporal resolution measures the rate at which images, from the same location, are retrieved with the same observation angle.

Presently there are multiple satellite systems collecting imagery (Rocchini et al. 2007; Toth et al.2016). Examples of these systems are Sentinel Mission, Landsat , the Moderate Resolution Imaging Spectroradiometer (MODIS), IKONOS, the Advanced Very High Resolution Radiometer (AVHRR), among others.

Satellite series like Sentinel are revolutionizing the sector by providing freely available high spatial, spectral and temporal resolution data, therefore, Sentinel 2 was the data source used this work, since it provides thirteen different multispecral bands, in three distinct spatial resolution (10 meter, 20 meter and 60 meter). More on this will be discussed in the next section.

2.1.2.2 SENTINEL Satellite images

The Sentinel missions are being developed by ESA (European Space Agency), specifically for the operational needs of the Copernicus programme (https://www.copernicus.eu/en/about-copernicus). Each Sentinel is based on a constellation of two satellites to guarantee revisits and coverage requirements, providing robust datasets for Copernicus Services. These missions carry a range of technologies, such as radar and multispectral imaging instruments for land, ocean and atmospheric monitoring. Currently, there are planned seven missions on board Copernicus, where in two, the data comes from satellites with the necessary multispectral information: Sentinel-1 and Sentinel-2.

Each Sentinel (1 e 2) provides the view of spectral data about the satellite images. Sentinel- 1 has three types of data products, which are Level-0 , Level-1 and Level-2. In Sentinel-2, data information is acquired on 13 spectral bands, in the visible, near-infrared (VNIR) and short-wavelength infrared (SWIR) spectrum In this work we focus on Sentinel 2 images because they have available several multispectral bands, with 13 bands in visible, near infrared and short-wave infrared part of spectrum. (Figure 2.5. depicts an example of image from Band 11).

Figure 2.5: Sentinel 2 Band 11 (grey area) in contrast with natural satellite image (Source: QGIS 3.4.5, imagery from year 2019)

Sentinel-2 mission has an almost-polar orbit and has a MultiSpectral Instrument (MSI)

sensor with thirteen multispectral bands with big spatial resolution. It includes a temporal resolution of 10 days with one satellite and 5 days with the two other operational satellites. It aims at monitoring variability in land surface conditions and supports monitoring of Earth's surface changes.

Moreover, its mission is a constellation with two twin satellites, Sentinel-2A and 2B, and has a spatial resolution of 10 m, 20 m and 60 m. Figures 2.6, 2.7 and 2.8 [1] show, for each spatial resolution, which band (wavelength) is available:

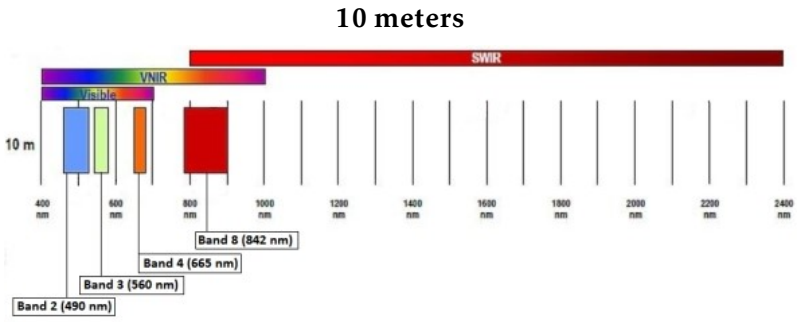


Figure 2.6: SENTINEL-2 10 m spatial resolution bands: B2 (490 nm), B3 (560 nm), B4 (665 nm) and B8 (842 nm).

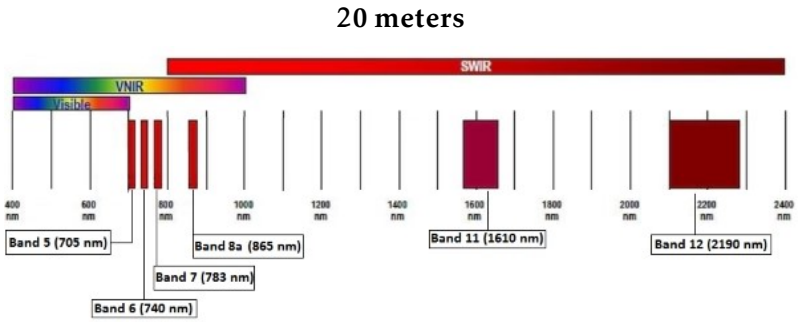


Figure 2.7: SENTINEL-2 20 m spatial resolution bands: B5 (705 nm), B6 (740 nm), B7 (783 nm), B8a (865 nm), B11 (1610 nm) and B12 (2190 nm) .

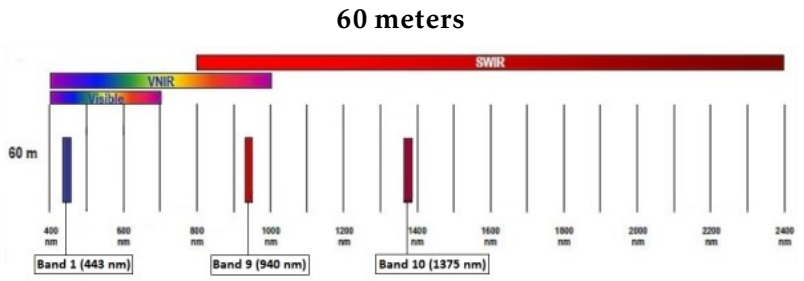


Figure 2.8: SENTINEL-2 60 m spatial resolution bands: B1 (443 nm), B9 (940 nm) and B10 (1375 nm).

[1] https://earth.esa.int/web/sentinel/user-guides/sentinel-2-msi/resolutions/spatia

As observed in Figures above, Sentinel 2 presents a large variety of multispectral bands and resolutions, when compared with other satellite systems. For example, Figure 2.9 depicts the comparison of Sentinel 2 and Landsat 8 wavelength cover of each one, and their spatial resolution levels.

	Sentinel-2A MSI			Landsat 8 OLI		
Band	Spectral region	Wavelength range (nm)	Resolution (m)	Spectral region region	Wavelength range (nm)	Resolution (m)
B1				Blue	435–451	30
B2	Blue	458–523	10	Blue	452–512	30
B3	Green peak	543–578	10	Green	533–590	30
B4	Red	650–680	10	Red	636–673	30
B5	Red edge	698–713	20	NIR	851–879	30
B6	Red edge	733–748	20	SWIR1	1566–1651	30
B7	Red edge	773–793	20	SWIR2	2107–2294	30
B8	NIR	785–899	10			
B8A	NIR narrow	855–875	20			
B11	SWIR	1565–1655	20			
B12	SWIR	2100–2280	20			

Figure 2.9: Spectral bands range and spatial resolution of Sentinel-2A MSI and Landsat 8 OLI sensors (Cerasoli et al. 2018).

Finally, it should be highlighted that Sentinel 2 guarantees the possibility of combining different bands, and therefore, to build vegetation indices (Bannari et al. 1995). Building vegetation indexes is of utmost importance for the development of this work – they allow distinguishing different types of land - and also the reason for selecting Sentinel-2 as the main data source for this dissertation.

2.1.3 Vegetation Indices

To accomplish the task of analyzing remote sensed images to identify and classify land (eucalyptus trees in this work), we need to characterize this type of tree, for instance using the canopy leaf index or biomass. These characteristics are known as Vegetation Indices (VI). There are a large set of vegetation indices, but for this work, we selected the most appropriate for fitting the desired output (described on section 5.1.1).

Vegetation indices are precise quantitative measurements that indicate the vigor of the vegetation (Campbell, 1987). For example, the results of vegetation indexes show better sensitivity when compared with results of individual spectral bands ,for the detection of biomass (Asrar et al. 1984).

Furthermore, these indices are extremely useful for the interpretation of remote sensing images. For instance, they establish a distinctly method for detection of land use changes (multitemporal data), crop prediction (Baret, 1986) and the evaluation of vegetative cover density. In the area of thematic mapping - objective of this work - these indices enable huge improvements of vegetation types classification (Asrar et al. 1984; Bariou et al. 1985; Qi et al. 1991; McNairn and Protz, 1993).

To calculate vegetation indices we need to pay attention to the electromagnet spectrum that composes the used bands, this is, its spectral signature (Section 2.1.1). Figure 2.10 depicts the electromagnet spectrum characteristics and for this study, the bands with more relevance are the ones covering the visible and near infrared.

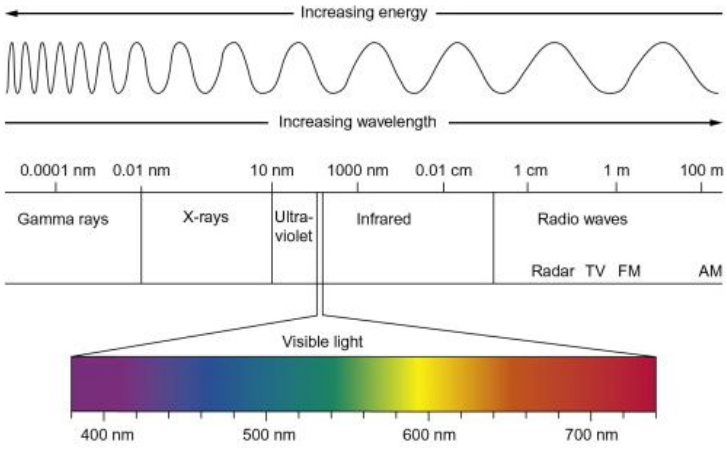


Figure 2.10: The electromagnetic spectrum (Michael F. L'Annunziata, 2016).

Understanding the most important interactions of radiant energy with the Earth's surface is very important for solving for remote sensing problems (Huete, 1989). Over the decades, some vegetation indices have shown very good interrelationship with distinctive factors of interest, like forest areas and their biomass content (Perry and Lautenschlager, 1984; Baret 1986). Pearson and Miller (1972) are pioneers in the history of vegetation indices. The first two indices were developed by them in the form of ratios: the Ratio Vegetation Index (RVI) and the Vegetation Index Number (VIN), for the reckoning and monitoring of vegetative covers.

$$\mathrm{RVI} = \mathrm{R/NIR}$$
$$\mathrm{VIN} = \mathrm{NIR/R}$$

Where R is the mean reflectance in the red channel and NIR is the mean reflectance in the near infrared channel. Through the years, a large set of vegetation indices were developed attending to a big diversity of land types (Pesant et al. 1995).

In land cover classification supported by remote sensing systems, the most important and widely used index is the Normalized Density Vegetation Index (NDVI) (Karnieli et al. 2010). NDVI is a simple but effective vegetation index for quantifying the vegetation (Rouse et al. 1973). All surfaces absorb and emit light energy in different frequencies and wavelengths, as seen before. The human eye, for example, can see the light strip between around 400 and 750 nm (Figure 2.10). Since vegetation can absorb and reflect a bigger light strip than us (Section 2.1.1), NDVI measures the state of health of the plant

in question, giving the result in an interval between -1 and 1 (See more details in section 5.1.1).

Figure 2.11: Levels of reflectance, regarding each health state of a leaf (Source: agricolus.com/en/indici-vegetazione-ndvi-ndmi-istruzioni-luso/).

The percentage of refracted radiation in specific bands, such as near infrared (NIR), red (Red), and short-wave infrared (SWIR), varies with plant health and water stress. The formulation of the NDVI index is:

$$NDVI = \frac{(NIR - Red)}{(NIR + Red)}$$

Which corresponds to the ratio between the difference and the sum of the refracted radiations in the near infrared (NIR) and in the red (Red).

Besides NDVI (most used), there are many other popular vegetation indices in remote sensing, such as: Normalized Difference Moisture Index (NDMI or NDWI) (Xu 2006); Leaf Area Index (LAI) (Asner et al. 2003); Enhanced Vegetation Index (EVI) (Matsushita et al. 2007); Green Normalized Difference Vegetation Index (GNDVI) (McFeeters et al. 1996); Soil Adjusted Vegetation Index (SAVI) (Gilabert et al. 2002); among others. Regarding this thesis work, we used some aforementioned indices (NDVI, NDMI and GNDVI) plus: Chlorophyll Index Green (CIG); Soil Composition Index (SIG); and Green Vegetation Moisture Index (GVMI). All the used indices will be described in Chapter 4.

A good summary about the chosen vegetation indices can be seen in (http://enviidl.com/help/Subsystems/envi/Content/Vegetation%20Analysis/BroadbandGreenness.htm).

2.1.4 Eucalyptus

Currently, eucalyptus´ trees plantation area is increasing in Portugal, in relation to other species, because it is the main supplier of raw material for cellulose industries and provides quick return investments (Garcia 2017). Eucalyptus trees represents a very important role in the Portuguese economy, not only by their importance regarding the size of occupied land area, as well as their profitability for industrial sectors, particularly the cellulose industry (Monteiro Alves et al. 2007). Figure 2.12 shows the distribution of forests´types in Portugal, where it can be observed that Eucalyptus represent 26% of country forest area, followed by Pine trees and cork trees with 23% each.

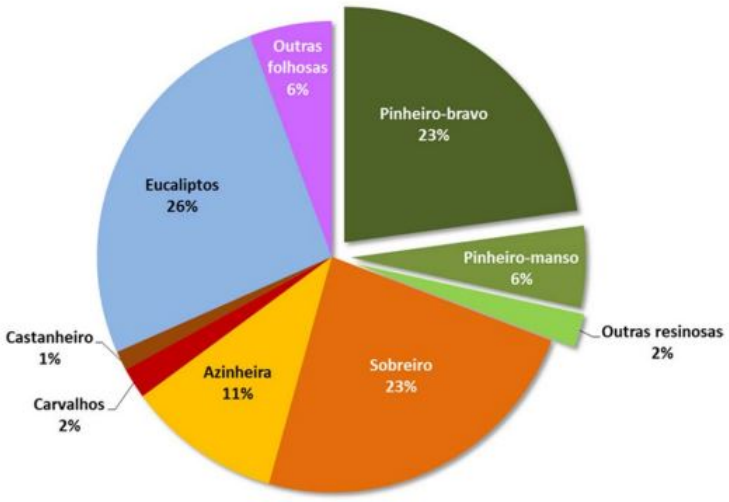


Figure 2.12: Distribution of total tree areas by specie/group of species (ICNF 2013).

Figure 2.12 shows the percentage of occupancy of the eucalyptus trees (26%), being the most common tree in Portugal.

Eucalyptus are a hardwood type of tree, originated from Australia, that is now very common in Portugal for paper production. This tree main distinctive characteristic is its height, which can reach 70-80 meters when adult (Naturelink 2009), which makes them one of the tallest trees in the world. Other interesting characteristics are: its wood is whitish; it is very rich in water (very humid); and it displays a "difficult"relationship with neighbouring vegetation because it absorbs most nutrients leaving nothing else for other tree types (Liu et al. 2010).

As a fast-growing species, eucalyptus have a relevant bio-climatic adaptation capacity. The amount of water in the soil determines, how tall Eucalyptus will grow and their degree of biomass abundance of biomass, as well as the woody matter of its stem (Alves et al. 2007). Another unmistakable characteristic of eucalyptus forests, is that the trees always appears very close to each other and grow in straight lines, giving the impression that were geometrically man-made.

Further, when this tree is inserted in a forest settlement, its trunk will be very tall and straight. Due to Eucalyptus´ trees special characteristics their effective identification is an interesting and important subject for monitoring land use changes, as well as investigation on environmental quality (Qiao et al. 2016).

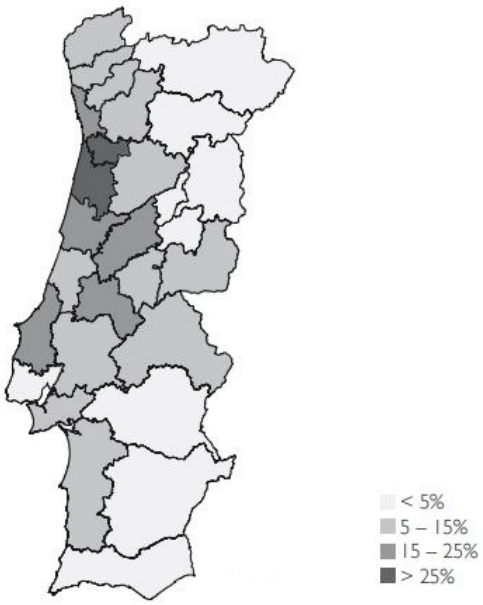


Figure 2.13: Eucalypt distribuiton in Portugal: proportion of the area occupied in each NUT III (National Forest Inventory 2005-06).

NUTS are regional divisions that exist in all member-states of European Union, used by Eurostat for development of all regional statistics and by the European Union regional policy definition and financing of the cohesion funds. There are three different NUTS (I,II and III), each one with its own number of divisions. Figure 2.35 depicts NUTS III division, where each area estimates the percentage of eucalyptus that exist there, based on the total number of eucalyptus in Portugal.

As mentioned before, to classify and identify Eucalyptus trees, this study used multi-spectral satellite images of a small region in Continental Portugal that is known to have eucalyptus trees. The main goal was to identify and distinguish them, from other types of trees, applying a data fusion approach.

2.1.5 Land Cover Classification Methods

Land cover classification methods goal is the development of interpreted maps, supported by remotely sensed images. Usually , classification is done by visual interpretation of features and the manual delineation of their boundaries (for example, COS). However, with advances on computer science and digital imagery and computer assisted interpretation the classification process to be much faster and consequently most classification works, today, use digital classification procedures (Eastman 2003).

The four most used approaches for land cover classification are: supervised classification, unsupervised classification (Eastman 2003), pixel-based classification and object-based classification. The first two (supervised and unsupervised classification) are more general and can be applied in many situations related with machine learning problems (Chaovalit et al. 2005). Pixel-based classification and object-based classification are more specific for land cover classification (Weih et al. 2010).

Supervised classification. Allows a supervised formalization of the way in which expected land cover classes should appear in the set of images (Eastman 2003). After, a procedure, named classifier is used to evaluate the likelihood that each pixel belongs to one class (Eastman 2003). The main logic behind supervised classification is the previous knowledge of the classifier system, i.e. there is a previous selection of representative samples for each land cover class, known as training points, which then are then applied to the entire image. Training sets, for each class, can be established using land visits, maps, air photographs or even photo-interpretation. The next step is to use the training data to estimate the parameters of the chosen classifier algorithm using methods such as Support Vector Machine, Maximum Likelihood Classifier, etc., to determine the properties of the probability model. The last step is the classification, using the previous trained classifier to classify every pixel in the image, assigning to each pixel the desired ground cover types (Richards et al. 1999).

Unsupervised classification. Contrary to supervised classification, this type of classifier is used to discover frequent occurrences and reflectance patterns in the set of images, supposing that these represent major land cover classes. Then, the identity of each class is found by a combination of experience or ground truth validation (for example, visiting the study area and observing the actual cover types). In other words, unsupervised classification does not require prior knowledge (Eastman 2003). I.e. it groups the data into the most frequent spectral groupings, usually called clusters, associated with the data. Then, the system's analyst identifies the resulting classes by joining a sample of pixels in each class with available ground truth data (Eastman 2003). The most commonly image clustering algorithms are K-means or ISODATA (Theiler et al. 1997; Phillips et al. 2002).

Pixel-based classification. This type of classification is done by attempting to classify every pixel of the original image, independently of each other. This is one of the most used approaches for land cover classification, mainly because the pixel is the fundamental unit of any satellite image. In land cover classification, some classes allow classifications constant and stable, like water bodies or roads, but when more details are needed, pixel-based techniques face problems such as assigning more than one class to a pixel, contributing to wrong estimation of land types (Foody 2002). There are several approaches used in pixel-based classification like the Maximum Likelihood Classifier (Yan et al. 2006; Myint et al. 2011) or Neural Networks (Kussul et al. 2017).

Object-based classification. The procedure involves categorization of pixels based on their spectral characteristics, such as texture, shape and/or spatial relationship with the surrounding pixels. The main difference of this approach from pixel-based classification is that object-based classification is based on information from a set of alike pixels, called objects, while pixel-based is based in each pixel. It starts by grouping the pixels in the current image into objects, a process named segmentation.

The objects are generated, with different scales, simultaneously in an image, representing features. Several bands can be used to create those objects. After segmentation into different objects, this technique classifies each object, instead of each pixel. A famous classifier used by this approach is the Nearest neighbor (NN) (Yu et al. 2006).

2.1.6 Data Fusion for Land Cover Classification

With improvements in the availability of multisensor image data, from the current Earth satellites, the fusion of image data is turning into a very useful tool for remote sensing image classification (Pohl et al. 1998; Schmitt et al. 2016; Ghamisi et al. 2018). This, is the basis for the data fusion approach used in this work. Next, we describe several Computational Intelligent Approaches that are being applied in data fusion.

2.1.6.1 Computational Intelligent Approaches

Recently, many Computational Intelligent (CI) approaches for information/data fusion were proposed for usage in remote sensing approaches (Fauvel et al. 2006; Ayhan et al. 2012), such as decision trees (DT), artificial neural networks (ANN) and fuzzy logic, among others (Friedl et al. 1997; Kanellopoulos et al. 1997; Dai et al. 1999; Pal et al. 2003; Fauvel et al. 2006; Mas et al. 2008; Hsu et al. 2009; Ross et al. 2010; Mora et al. 2013; Ahmed et al. 2015; Colditz et al 2015; Santos et al. 2016; Schmitt et al. 2016; Mora et al. 2017).

A very interesting study compared three classifiers: Support Vector Machines (SVM), Gaussian maximum likelihood classifier (applied the leave one-out-covariance procedure (GML-LOOC)) and k-Nearest Neighbor (*k*-NN). This study consisted on fusion of hyperspectral and LIDAR remote sensed data for classification of non-trivial forest areas (Dalponte et al. 2008). The results suggested that SVM classifier provided much better results (bigger accuracies) than the other two classifiers, and *k*-NN technique proved unsuitable for solution of hyperdimensional problems.

Although the classical Computational Intelligent Classifiers, described above, be the most widely used procedures in remote sensing, there are other approaches that are increasing their influence on data fusion of remote sensing problems. These approaches are regarding fuzzy logic, known as Fuzzy Inference Systems (FIS) (Ross et al. 2010). Example of the application of this technique appeared on an experiment that the output of it were the results of two procedures, both of them classified with pixel-based technique, that were compared. The conclusion of the study was that, considering the level of classification accuracy, fuzzy logic can be satisfactory used for image classification, therefore, in remote sensing problems. (Nedeljkovic 2004)

Although Computational Intelligent Classifiers based on decision trees and ANN are the most widely used in remote sensing, there are other approaches based on fuzzy logic have been applied to remote sensing and data fusion (Mora et al 2013), (Mora et al 2017), (Schmitt et al 2016), (Santos et al 2016). (Nedeljkovic 2004). Data fusion based on fuzzy logic techniques is emerging as a useful technique for land classification, particularly through enhanced aggregation operators to perform correct reasoning inferences (Hyder et al. 2012; Santos et al. 2016). The aforementioned aggregation operators have been applied to decision-making methods as well as data fusion, to ensure positive or negative reinforcement in the classification results (Beliakov et al. 2007; Torra et al. 2007; Ribeiro

et al. 2010; Rudas et al. 2013). Aggregation operators will be further discussed on section 3.1.3.

With the objective of discussing the suitability of different CI methods for studying land cover spatiotemporal modifications, mostly for improving land usage and monitoring, there is an interesting study where one fuzzy image fusion technique was compared with Decision Trees and ANN methods, for fusing images (Mora et al. 2015) (Mora et al. 2017). This study consisted on an extension of previous work where a fuzzy-fusion approach with reinforcement aggregator operators for land cover classification from multispectral satellite images was tested. Further, the work on this thesis is based on this latter algorithm for data fusion, but instead of being applied to classification of several classes we only applied it to one class, Eucalyptus. The study revealed that the produced results with the fuzzy fusion approach are adaptable and consistent with DT (similar), but having higher disparity in the classification of grassland and shrubland (Mora et al. 2017).

Decision Trees

It is a well-known classification method that consists in partitioning the input data into smaller subsets with some similar features, with the goal of turning them homogeneous, defining a set of rules to be followed sequentially down an inverted tree-like structure (Pal et al. 2003). DT are composed by nodes and leaves, where each node represents a rule applied to the data. Root node is at the top, followed by a set of internal splits until it reaches the terminal node. Each node has only one parent (its origin) node and one (or more) descendant nodes. Terminal nodes have no descendant nodes. About the classification procedure, the input is introduced into the designed tree and the set of defined rules will determine the path that the input should follow, starting at the root, ending at one terminal node.

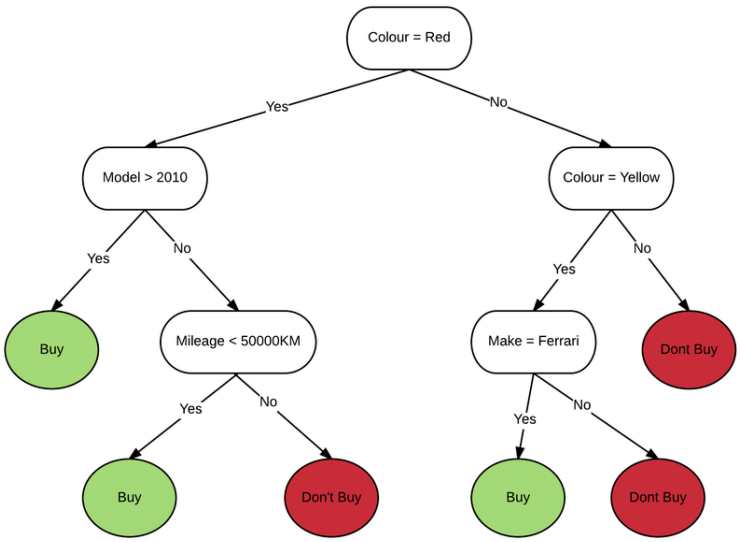


Figure 2.14: Example of a decision tree (Moodley 2016).

Figure 2.15 depicts a decision tree, with the top node being the root node, the other white nodes are the internal split nodes and the colored ones are terminal nodes.

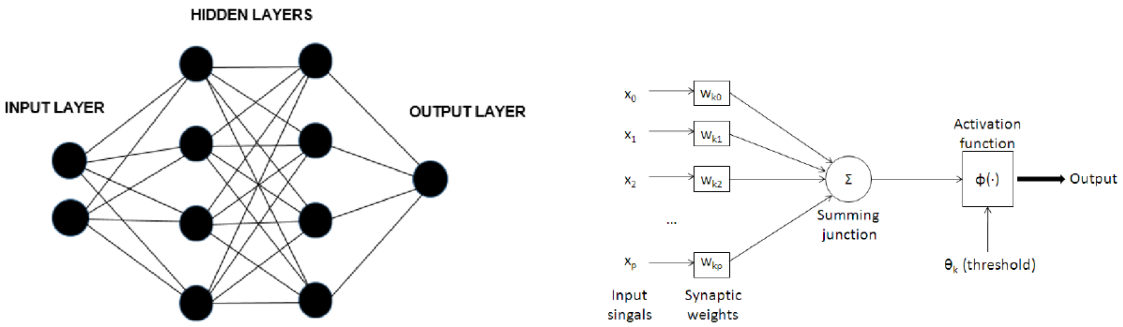
About DT design, it can be done in two ways:

- Manually, when based on a priori knowledge;
- Automatically, using mathematical evaluation algorithms (ID3, C4.5 or CART)

DT have been applied in remote sensing land classification since the 90´s (Friedl et al. 1997; Xu et al. 2005; Fauvel et al. 2006; Colditz et al. 2015; Mora et al. 2017). Particularly, Piiroinen used a DT approach for identification of a specific species of eucalyptus tree (Piiroinen et al. 2017).

Artificial Neural Networks

ANN is a classification technique based on a cluster of neurons where each connection can transmit signals from one neuron to another. Neurons are trained to detect patterns in the training data set and then the trained classifier is applied to unknown data. ANN can be multi-layer, where the number of artificial neurons in the input and output is determined by the data being analyzed, whereas the number of hidden layers is determined principally by trial and error (Ayhan et al. 2012).



(a) ANN organization

(b) Mathematical model of an ANN

Figure 2.15: a) The organization of an ANN (Moodley 2016) and (b) mathematical diagram of an ANN (Martínez-Álvarez et al. 2015).

The main objective behind ANN is to identify a set of weights that minimize the sum squared errors (SSE) of all predictions made by the ANN. To accomplish this task, there are some algorithms to train the network, such as the Levenberg-Marquardt, quasi-Newton, back-propagation, among others (Mukherjee et al. 2012).

Examples of land cover classification with ANN can be seen in (Mora et al. 2017; Canziani et al. 2008; Mas et al. 2008; Yang et al. 2018).

Fuzzy Inference Systems

This method is based on rule models, defined with logic operators, where the rules establish relationships between fuzzy sets. For example, IF-THEN rule is a basic fuzzy logical rule, which means:

IF < fuzzy proposition > THEN < fuzzy proposition >,

where fuzzy are placed before and after the THEN statement, giving sense to the concrete problem. An example is,

IF temperature > 30 THEN fan speed is 3,

this means that when the fuzzy system checks the value of the temperature, if the condition is true (is indeed > 30), the consequent proposition will be processed (i.e. the fan will be set to 3. This example is a basic logical rule model, with only one operator (>). Fuzzy proposition antecedents can be evaluated using other types of operators, like intersection or union (more details in section 3.1.1).

FIS have three major procedures (Ross et al. 2010):

1. Inputs fuzzification
2. Fuzzy rules;
3. Inference scheme selection.

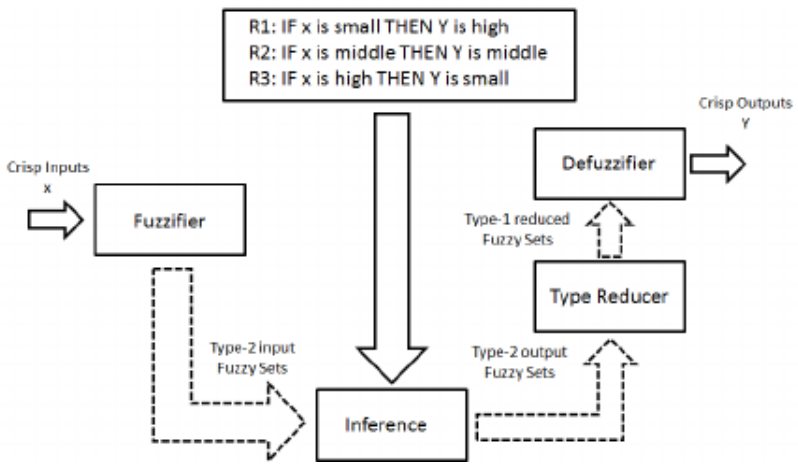


Figure 2.16: Fuzzy Inference System diagram (Castañón-Puga et al. 2013).

The most well-known models for the inference scheme are those of Mamdani (1975) and Takagi-Sugeno (1985). Mandani scheme transforms a fuzzy input into a fuzzy output with the max-min operator and Takagi-Sugeno transforms a fuzzy input into a crisp

output with a weighted average scheme. Mandani output needs de-fuzzification, which is accomplished with techniques such as center-of-area (Ross 2010).

In summary, CI techniques applied to data/information fusion are proving very useful tools in terms of decision-making models for both selection or classification problems in remote sensing (Hyder et al. 2002; Bleiholder et al. 2008; Lee et al. 2010) (Mora et al 2015). As mentioned before, another interesting algorithm for dealing with data fusion issues is called Fuzzy Information Fusion (FIF) (Ribeiro et al 2014). This algorithm is based on fuzzy logic and specialized decision-making aggregation operators and was applied to Spacecraft landing with hazard avoidance (Bourdarias et al 2010) and Land coverage classification (Mora et al 2013: Mora et al 2015). This is the chosen algorithm to be used in this work, for the data fusion of satellite images, with the aim of identifying Eucalyptus trees. The details of this algorithm are presented in Section 3.1.2.

In the next chapter can be seen with more detail Fuzzy Logic concepts.

Chapter 3

Data Fusion Approach

This chapter presents the main background concepts underlying the proposed approach, more concretely, data fusion.

Data fusion consists of a process of aggregating data from various sources to construct a single compound of all data received, with higher quality of information (Bleiholder et al. 2009; Lee et al. 2010; Hyder et al. 2012). This process includes three main types: image fusion, multisensor fusion and information fusion (Hyder et al. 2012) (Ribeiro et al 2015). The sources to be fused (e.g. images or information) all have in common the fact that they must regard the same subject.

Image fusion. The main objective of image fusion is to decrease uncertainty and redundancy, maximizing pertinent information by merging several image representations of the same scene (Goshtasby et al. 2007). The respective algorithms are usually divided into two groups: pixel based and feature-symbolic based (Dai et al. 1999). In the first case, the most common, data is fused pixel-by-pixel, while the second case requires extraction and fusion of features from different sources (Piella 2003; Hsu et al. 2009).

Multisensor fusion. This method refers to fusion of data, acquired from sensors, and its main goal is to gather data measurements, extracted from different sensors, and combine them into a single representation (Ribeiro et al. 2014). Most common techniques use statistical methods (e.g. Kalman filters) and probabilistic techniques (e.g.Bayesian networks) (Waltz et al. 1990; Manyika et al. 1995; Klein 2004; Lee et al. 2010; Goodman et al. 2013). Further, hybrid strategies join different multisensor fusion techniques, taking advantage of their individual benefits and dropping some flaws.

Information fusion. Information fusion is a multi-level process of combining different data to produce fused information (Torra et al. 2007; Lee et al.2010). There is a tenuous line between image fusion and information fusion, for example, feature and

symbolic levels of fusion are sometimes considered image fusion but they can also be considered as information-based fusion (Piella 2003; O'Brien et al. 2004).

In general, data fusion approaches, based on fuzzy logic techniques, are emerging as a good technique for land classification to perform correct reasoning inferences (Hyder et al. 2012; Santos et al. 2016). An interesting algorithm for dealing with data fusion issues is called Fuzzy Information Fusion (FIF) (Ribeiro et al 2014). As mentioned, this algorithm is based on fuzzy logic and specialized decision-making aggregation operators and was applied to spacecraft landing with hazard avoidance (Bourdarias et al 2010) and Land coverage classification (Mora et al 2013: Mora et al 2015). Because of its versatility, FIF is the chosen algorithm to be used in this work for identifying eucalyptus trees.

Next will be presented some concepts regarding fuzzy logic and its components.

3.1 Fuzzy Logic

This CI technique is very important for data fusion processes (Ribeiro et al. 2014). The term fuzzy, coined by (Zadeh, 1988), refers to concepts or variables that are not clear or sometimes they are vague. It is common to face situations whether we can't determine if the state of a problem is true or false,, so fuzzy logic can provide a valuable mechanism for reasoning under uncertainty. In this section we will focus on the fuzzification process because it is the only part used in the data fusion approach. Fuzzification allows normalization of data sets by means of membership functions, thus transforming any measurement (or image) into a normalized [0-1] scale, allowing degrees of uncertainty. The interested reader can see further details in (Ross et al. 2010).

The gradual evolution of the expression of uncertainty using probability theory was challenged, first in 1937 by Max Black (studies in vagueness), then with the introduction of fuzzy sets by Zadeh (1965) (Ross, 2010). In another Zadeh's paper (1995), he challenged not only probability theory as the only representation of uncertainty, but also the very foundations upon which probability theory is based: classical binary (two-valued) logic (Klir and Yuan, 1995).

Over the years, several authors have shown and proven that fuzzy systems are quasi-universal approximators (Kosko, 1994 ; Ying et al. 1999). A Fuzzy Logic System has several advantages:, a) works with any type of input, whether it is imprecise, distorted or noisy input information; b) its construction is easy and understandable; c) it provides a very efficient solution to complex problems in many fields, as it resembles human reasoning and decision making processes.

3.1.1 Fuzzy Sets and Membership Functions

In classical (crisp) sets the transition of an element in the universe between membership and non-membership to a given set is abrupt, either belongs or not to the set (usually called "crisp") (Ross 2010). In the fuzzy universe, an element belongs to that set, in a

gradual form. The degrees of membership can be viewed as conforming to the fact that the boundaries of the fuzzy sets are vague and ambiguous. A fuzzy set and its elements are measured by a function that attempts to describe its vagueness and ambiguity (Ross 2010). Formally, a fuzzy set A, in a universe of discourse U, is characterized by a membership function $\mu_A(x)$->[0,1], that represents the degree of membership of x in A (Zadeh, 1995),

$$\underset{\sim}{A} = \left\{ \frac{\mu_{\underset{\sim}{A}}(x_1)}{x_1} + \frac{\mu_{\underset{\sim}{A}}(x_2)}{x_2} + \cdots \right\} = \left\{ \sum_i \frac{\mu_{\underset{\sim}{A}}(x_i)}{x_i} \right\}.$$

Figure 3.1: Fuzzy set universe discrete and finite.

$$\underset{\sim}{A} = \left\{ \int \frac{\mu_{\underset{\sim}{A}}(x)}{x} \right\}.$$

Figure 3.2: Fuzzy set universe continuous and infinite.

Figure 3.1 depicts the convention for fuzzy sets when the universe of discourse, U, is discrete and finite for a fuzzy set A, while in Figure 3.2 is continuous and infinite (Ross 2010).

As an example, let's consider the degree of membership of 15ºC for the fuzzy sets cold, mild and hot: $\mu_{cold}(15)=0.5$, $\mu_{mild}(15)=0.75$ and $\mu_{hot}(15)=0.2$, where x is the temperature in degrees Celsius. From this example, it is possible to note that fuzzy sets can overlap.

Membership Functions. The characterization of a fuzzy set is given by its membership functions. So, there are several topologies for defining functions that better match each variable of a given problem (Ross, 2010).

A membership function of a fuzzy set A on the universe of discourse U is characterized by μ_A:U -> [0,1], where to each element of the universe U is assigned a value between the interval [0,1], called membership value (Zadeh 1965). It evaluates the degree of membership of the element in U to the fuzzy set A. The membership function can be represented graphically, where the x axis represents the elements of the universe and the y axis the corresponding membership value.

Since the information confined in a fuzzy set is described by its membership function, it is useful to develop a lexicon of terms to describe various special features of this function (Ross 2010). These terms are depicted in Figure 3.3.

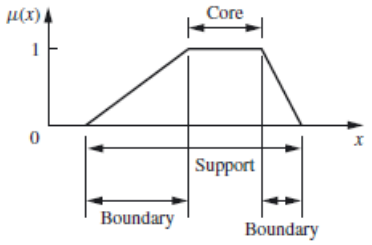


Figure 3.3: Terms defining membership functions (Ross 2010).

Core. For some fuzzy set A, the core is defined as a region of the universe with complete and full membership in the set A, i.e., the region that has the maximum value of the membership function (1)..

Support. For some fuzzy set A, support is defined as a universe region characterized by nonzero membership in the set A, i.e., the region where the membership function never reaches value zero.

Boundaries. For some fuzzy set A, its boundary is defined as a universe region containing elements that have a nonzero membership but do not have full membership in the set A, this is, respects the condition $0 < \mu_{\mathrm{A}}(x) < 1$.

There are three types of common functions: triangular membership function, Gaussian membership function and trapezoidal membership function. Sigmoid is commonly used in many situations (Zhao et al 2002). Graphical examples are depicted in Figure 3.4, 3.5, 3.6 and 3.7.

Triangular membership function. To define a triangular function it is necessary to assign three values, a lower limit **b**, upper limit **a** and a value **c** (top vertex), where the condition (b < c < a) must be respected.

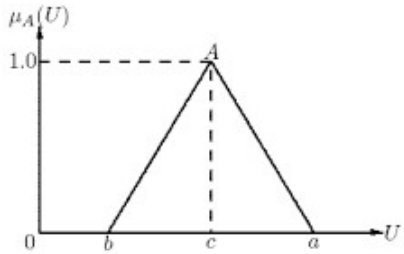


Figure 3.4: Triangular membership function (Jamsandekar et al. 2013).

Gaussian membership function. It is defined by a central value **c** and a standard deviation $\sigma > 0$.The smaller is the standard deviation, narrower is the function. The width of the function is defined by the values **a** (**c** - σ) and **b** (**c** + σ).

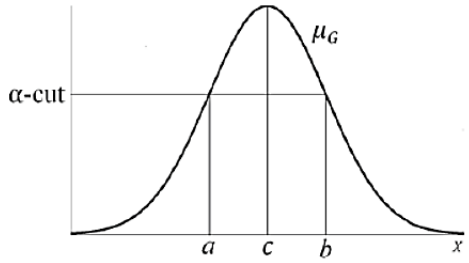


Figure 3.5: Gaussian membership function (Mohamad et al. 2016).

Trapezoidal membership function. Defined by a lower limit **a**, an upper limit **d**, a lower support limit **b** and an upper support limit **c**, where the condition (a < b < c < d) must be respected. This formulation gives to the elements between the interval [b,c] maximum membership value, this is, the value 1.

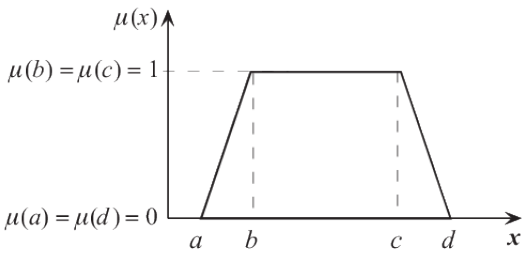


Figure 3.6: Trapezoidal membership function (Soliman et al. 2006)

Sigmoid membership function. A sigmoidal membership function is defined by two parameters: **m**, that controls its center value, and width, that controls the width of the function, influencing the membership values. A positive value of **m** means the left side approaches 0 while the right side approaches 1. It is quite appropriate for representing concepts such as "very large" or "very negative" (Dubois et al. 2012).

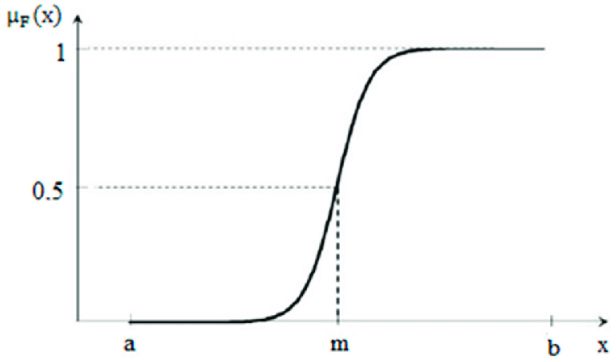


Figure 3.7: Sigmoid membership function (Soliman et al. 2006).

3.1.1.1 Fuzzification

Fuzzification is the process of transforming a set of data into a fuzzy membership function. It begins by choosing a suitable membership functions for representing the data set. There is the inverse process too, called defuzzification (Zadeh 1988).

There are numerous ways to define membership values (Ross 2010):

Intuition. This one is derived from experts to develop membership functions through knowledge andunderstanding of the context.

Inference. This method consists in the usage of expert knowledge to preform deductive reasoning, this is, we wish to deduce or infer a conclusion, given a body of facts and knowledge (form of rules).

Rank Ordering. This method accesses preferences by a single individual, a committee, a poll or other kind of opinion and can be used by assigning membership values to a fuzzy variable.

Inductive Reasoning. Here membership functions are generated exclusively based on available data.

In the context of this work, the more suitable for the development of membership functions was inductive reasoning, since the membership functions were built according to the available data (vegetation indices).

Fuzzy Sets Operations. By definition, a fuzzy set operation is any operation performed on fuzzy sets and are a generalization of crisp set operations. The most used operations are known as standard fuzzy set operations. There are three major operations that can be performed with fuzzy sets, union, intersection and complement (Zadeh 1965).

In other words, these three basic operators are used to build the definitions of the aforementioned standard fuzzy sets: standard union; standard intersection and standard complement (Zadeh 1965).

Standard union

As the basic union definition, when applied to two fuzzy sets (A and B), the output will be a set containing elements that belong to A or B. Union is associated to the MAX function:

$$\mu_{A\cup B}(u) = \max\{\mu_A(u), \mu_B(u)\}$$

Standard intersection

When applied to two fuzzy sets (A and B), the output will be a set containing elements that belong to A and B. Intersection is associated to the MIN function:

$$\mu_{A\cap B}(u) = \min\{\mu_A(u), \mu_B(u)\}$$

Standard complement

As the basic union definition, when applied to a fuzzy set (A), the output will be a set containing elements that not belong to A:

$$\mu_{\neg A}(u) = 1 - \mu_A(u)$$

Besides these most common operators, standard fuzzy set operators, there are other operators that can be used. Examples of those are the aggregation operators, which are mathematical functions that are used to combine data (Calvo et al. 2012). Among these aggregation operators, the most used belong to the classes of t-norms, t-conorms and; averaging operators. The full-reinforcement operators are less common but are quite versatile for penalizing or rewarding good or bad solutions (Ribeiro et al 2010).

T-norms

This family of operators are usually associated with the min operator and logical AND (standard intersection). There are many families of related t-norms that are defined by explicit formulas depending on some parameter. The main families of t-norms are Hamacher's (Zimmermann et al. 2011) , Yager's (Yager 1980) and Sugeno–Weber's (Weber 1983), as well as the well known min operator.

T-conorms

This family of operators, also called S-norms, refer to disjunctive functions that are usually associated with the max operator and logical OR (standard union). Like t-norms, there are many families of t-conorms, defined by explicit formulas depending on some parameter, such as Hamacher's (Zimmermann et al. 2011) , Yager's (Yager 1980) and Sugeno–Weber's (Weber 1983), as well as the well known max operator.

Averaging operators

This set of operators are usually associated with the mean operator. Unlike t-norms and t-conorms, averaging operators encompass a centering property (Yager et al. 1998): Ordered Weighted Averaging (OWA) is a well-known averaging operator, which permits to set weights to each score, before using the averaging function (Yager et al. 1998). More details can be seen in (Yager et al. 1998; Greco et al. 2010). Figure 3.8 depicts a plot of the OWA operator but in an academic context. Note that when dealing with fuzzy sets, the value interval is always between 0 and 1.

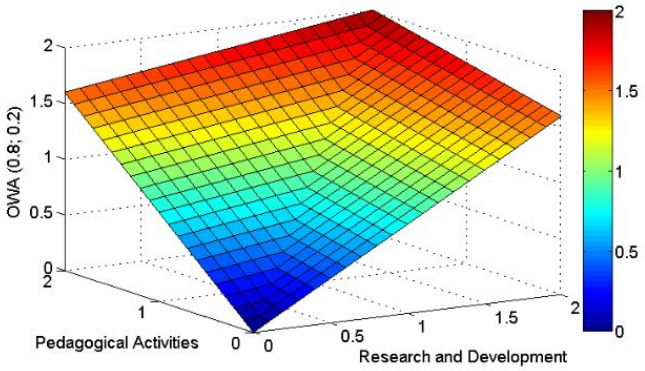


Figure 3.8: OWA applied in different context (Stoklasa et al. 2011).

Averaging operators are probably the most commonly used aggregation functions. An aggregation function f is averaging if, for every x, $min(x) \leq f(x) \leq max(x)$ (Campanella et al. 2011).

Full-Reinforcement operators

This kind of aggregation functions allows better modulation of the response because it penalizes “bad” results and rewards “good” ones. For example, in some cases (like the proposed work) it is useful to consider a positive (or upward) reinforcement behavior for high input values and a negative (or downward) reinforcement behavior for low ones. For instance, lets say we have a threshold of x=0.6, if all values in the aggregated score are less than x, (e.g. 0.3,0.3,0.5,), obviously we want to "punish"that result and keep the final aggregated score below 0.3 (downward reinforcement). Conversely, if all rates are above x, for example 0.8,0.8, we want to "reward"the combined result should be higher than 0.8 (upward reinforcement). Furthermore, when an operator is capable to perform upward and downward reinforcement it is called a full-reinforcement operator (Yager et al. 1998). All of the aforementioned operators (t-norms, t-conorms and averaging ones) do not display this behavior.

The two most used full reinforcement operators are FIMICA and UNINORM (Ribeiro et al. 2010).

Uninorm

Exhibits a conjunctive behavior, when in presence of low input values (below a certain neutral element $e \in]0,1[$), a disjunctive behavior for values higher than e and an averaging behavior otherwise (Campanella et al. 2011). In general an Uninorm operator is lies between t-norms and t-conorms, therefore, their properties are maintained (Yager et al. 1998).

An example of Uninorm operator U (Campanella et al. 2011) is:

$$U(x,y) = \begin{cases} e\ T_U\left(\frac{x}{e},\frac{y}{e}\right) & (x,y) \in [0,e]^2 \\ e + (1-e)\ S_U\left(\frac{x-e}{1-e},\frac{y-e}{1-e}\right) & (x,y) \in [e,1]^2 \end{cases}.$$

With the two squares (expressed in the equation), the role of uninorms in the rest of the unit square is not limited to each specific class of averaging functions, what lead to the characterization of several classes of uninorms (Campanella et al. 2011). Graphically, the UNINORM function is shown in Figure 3.9

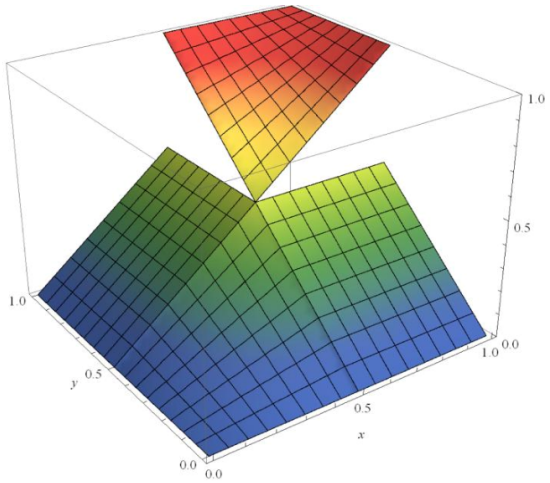


Figure 3.9: Uninorm operator with $e = 0.5$ (Bardalho et al. 2015).

To sum up, these type of functions gives more importance to relevant values and poor importance to lower values regarding the final output (Ribeiro et al. 2010). Since FIMICA is a full-reinforcement operator (it is an extension of UNINORM), used in this work, it is further described in Section 3.1.6, where all used aggregation operators are detailed.

3.1.2 FIF algorithm

The algorithm tested in this work is called Fuzzy Information Fusion (FIF) (Ribeiro et al. 2014) and combines intelligent computing (CI), specifically fuzzy logic, and multi-criteria decision-making methods. It is based on decision matrices and provides a method for merging heterogenous sources, producing a single component, to classify different alternatives. It is possible to infer that the FIF algorithm is a general algorithm (Ribeiro et al. 2014) that can be applied to any classification/selection problems, provided that the decision criterion can be formalized by fuzzy sets, representing any semantic concept (e.g. "low-slope"). It is also noteworthy that the FIF algorithm results from a work/study financed by ESA (European Space Agency), to test the safe landing of spaceships on planets (Bourdarias et al. 2010) and was already partially applied to specific remote sensing problems (Mora et al 2013; Mora et al 2015).

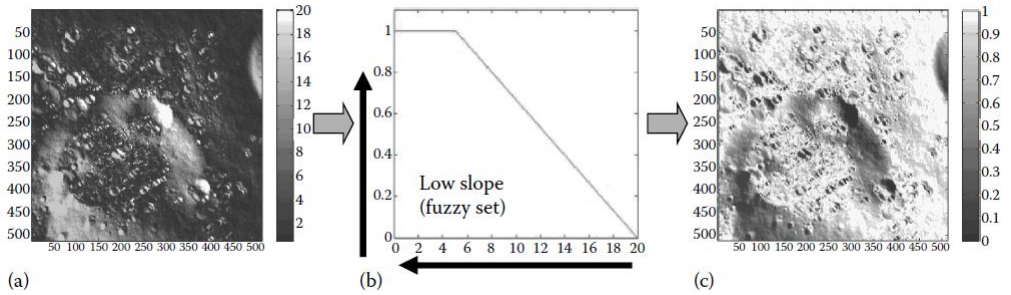


Figure 3.10: Example of normalization (fuzzification) of criteria in low-slope semantic on hazard maps. (a) Original hazard map. (b) Membership function topology. (c) Normalized map. (Mora et al. 2015)

The FIF algorithm includes four main steps:

1. Normalization process that includes a mathematical transformation of maps to ensure numerical and comparable information for aggregation (fuzzification);

2. Filtering of uncertain information regarding uncertainties and lack of confidence in the input data;

3. Setting a relative importance to each criterion , which depends on the satisfaction of the criterion for a specific alternative;

4. Choosing the aggregation method/operator to combine all processed data into a single compound.

The FIF algorithm starts (step 1) by running a customized data transformation process, for each of the input data sources to achieve the multisource data fusion. Figure 3.11 shows the algorithm architecture (Ribeiro et al. 2014).

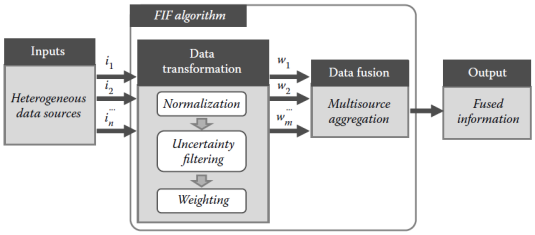


Figure 3.11: Fuzzy Information Fusion algorithm architecture (Mora et al. 2015).

Step 1: Normalization with membership functions (Fuzzification)

In this step it is considered that input data are gathered from heterogeneous data sources and those inputs are normalized using fuzzy membership functions, usually called fuzzification process (Ross et al. 2010). With this step, it is guaranteed not only normalized and comparable data but also representing data with semantic concepts (low slope or low-variance texture, for example), which simplifies problem understanding.

As mentioned before the challenge of normalization/fuzzification is to choose the best topology for the membership functions, because they depend on the context objective.

Step 2: Filtering uncertainty

The second step corresponds to dealing with any uncertainty found in the input data to be fused. This step is crucial when in presence of doubtful and imprecise acquired data from sensors since the uncertainty will affect all input values. This process is made by a filtering function (Figure 3.12). This function enables to adjust the membership functions in way to reflect the embedded uncertainty regarding input information and also to incorporate a pessimist or optimist view. Figure 3.12 also depicts the definition of the accuracy interval.

$$fu_{ij} = wc_j{}^*(1 - \max_{x\in[a;b]}\{|\mu(x) - \mu(x_{ij})|\})^*\mu(x_{ij})$$

$$a = \begin{cases} \min(D) & \text{if } x_{ij} - a_{ij} \le \min(D) \\ x_{ij} - a_{ij} & \text{if } x_{ij} - a_{ij} > \min(\boldsymbol{D}) \end{cases}$$

$$b = \begin{cases} x_{ij} + a_{ij} & \text{if } x_{ij} + a_{ij} \le \max(D) \\ \max(D) & \text{if } x_{ij} + a_{ij} > \max(\boldsymbol{D}) \end{cases}$$

Figure 3.12: Filtering function (left) and inaccuracy interval formalization (right).

Where:

w_{cj} is the confidence associated to criterion j; x_{ij} is the value of j_{th} criterion for alternative i; μ (.) is the membership degree in a fuzzy set; The interval [a,b] for x is defined with a_{ij} (accuracy associated to criterion j for alternative i); D is the variable domain for any matrix at each interaction.

Further, the domains for the parameters are: Accuracy: [0,10]; Confidence (0-100%). Even though its name, accuracy value a_{ij} represents a deviation from a central value, indicating the amount of inaccuracy in the observations ($[x_{ij} - a_{ij}, x_{ij} + a_{ij}]$).

In Figure 3.13 we show an example of filtering a fuzzy sigmoid function representing the concept "low texture" in soil, using accuracy of 5 and a 60% confidence. It is obvious how reduced the filtered values become and that the bigger distances from central point mean higher penalties $\mu(x)$, hence, the final fusion result will reflect our mistrust on the input data.

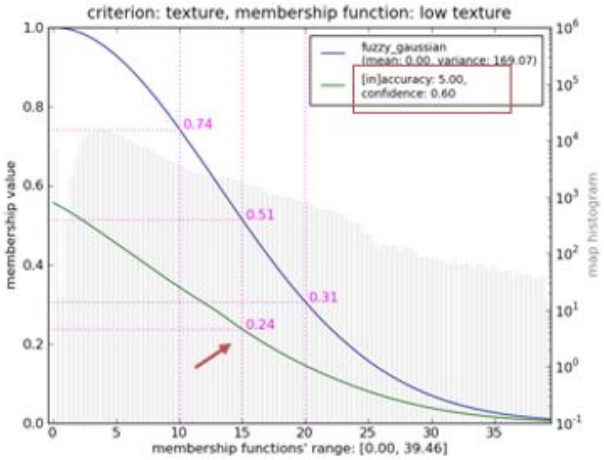


Figure 3.13: Illustrative example of filtering uncertainty (Ribeiro et al 2015).

Following this step, the input values that include any type of uncertainty can be taken in consideration without loss of robustness of the whole system and the parameters can be customized for any information fusion problem. In the context of this thesis since we use the only available satellite images, we do not use this step in our approach.

Step 3: Relative Importance of criteria with weighting functions

To express relative importance of criteria, the FIF algorithm uses linear weighting functions (Ribeiro et al. 2003), the logic behind these functions is that the satisfaction value of a criterion of these functions must influence its predefined relative importance.

$$L(fu_{ij}) = \alpha \frac{1+\beta fu_{ij}}{1+\beta}, \text{where } 0 \leq \alpha \leq 1 \text{ and } 0 \leq \beta \leq 1$$

Figure 3.14: Modified linear function, *L(x)* , used on FIF algorithm (Ribeiro et al. 2014)

α - Expresses the relative importance of the different criteria;

β - Controls the ratio L(1)/L(0) = 1 + β between the maximum and minimum values of the effective generating function.

$\alpha_j, \beta_j \in [0, 1]$;

fu_{ij} - Accuracy and Confidence membership value for the j_{th} criterion of alternative j (given by the equation presented on Figure 3.12);

[a, b] - Inaccuracy deviation interval (Figure 3.12);

The equation presented on Figure 3.14 is used by FIF algorithm to increase the computational efficiency and understandability. Analyzing it, we conclude that when β = 0 it falls in the classical weighted average aggregation operator, this is, the weighting function does not depend on the criteria satisfaction.

α **(Criteria importance)**	β **(Slope decrease)**
Very important (VI = 1)	High (H = 1)
Important (I = 0.8)	Medium (M = 0.67)
Average Importance (A=0.6)	Low (L = 0.33)
Low Importance (L = 0.3)	Null (N = 0.0)
Very Low Importance (VL = 0.1)	
Ignored (Ig = 0.0)	

Table 3.1: Semantic weights and corresponding values for α and β (Ribeiro et al. 2014).

Parameters α and β gives the definition of the weighting functions morphology. Based on Table 3.1, α provides the semantics for the weighting functions (importance), while β parameter provides the required slope for the weighting functions, allowing penalization or rewarding. The usage of the weighting functions comes with the need to reward or penalize input criteria depending on the satisfaction of criteria. Figure 3.15 represents the relative importance of criteria graphically.

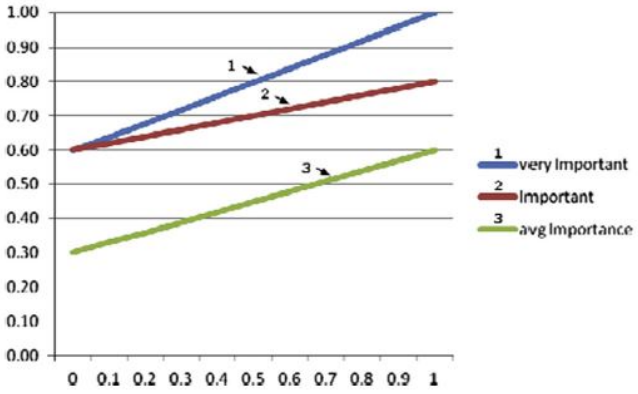


Figure 3.15: Linear weighting functions, l(x), for establishing relative importance of criteria VI, I and AVG using slopes (medium, low, high). (Ribeiro et al. 2014)

Step 4: Fusion process (Aggregation)

The fusion process is the last step of FIF algorithm and consists in aggregating/fusing all multisource input data processed in the previous steps. This aggregation is based on the combination of weighting functions (Ribeiro et al. 2003) with operators. Formally:

$$r_i = \oplus\,(W(fu_{i1}) \otimes ac_{i1}, \ldots, W(fu_{in}) \otimes fu_{in})$$

Figure 3.16: Formulation of the aggregation function (Ribeiro et al. 2003).

$\oplus$ - Aggregation method (e.g., sum, max, parametric operators);

$\otimes$ - Conjunction operator (e.g., multiplication, min);

fu_{ij} - Accuracy and Confidence membership value for the j_{th} criterion of alternative j (given by the equation presented on Figure 3.12);

$W(fu_{ij}) = \dfrac{(L(fu_{ij})}{\sum_{k=1}^{n} (L(fu_{ij}))}$, where $L(fu_{ij})$ is the weighting function depicted above (Figure 3.14).

Finally, since in this study, since we were only concerned with data fusion of satellite images (Note: not heterogenous sources, since it is the same image seen in different multi-spectral bands), step 2 (*Filtering uncertainty*)was not necessary because the confidence on the different bands was identical. In summary, our data fusion approach follows FIF for normalization of the images (vegetation indices) and then for aggregation we use several aggregation operators to construct an added-value single map, from where the classification of the eucalyptus was inferred.

3.1.3 Aggregation Operators

Since aggregation operators can significantly alter the results in combining information (Ribeiro et al. 2010; Rudas et al. 2013) they won relative importance upon the image fusion process.

Aggregation operators have been widely studied and developed and their usage in fuzzy multicriteria problems is broadly spread (Beliakov et al. 2007; Calvo et al. 2012). When we are facing an information fusion process, the right aggregation operator must also be carefully considered, because it is a main issue in this kind of problems (Beliakov et al. 2001).

There are many famous classical aggregation operators, but the most well-known are: max-min methods, generalized mean methods (weighted sum and product), outranking methods, distance-based methods and pairwise comparison methods (Calvo et al 2012; Triantaphyllou 2000; Tzeng et al. 2011). Reinforcement operators are relatively unknown, but they are the only ones that can guarantee either positive or negative reinforcement when fusing values (Yager et al. 1998; Beliakov et al 2007; Ribeiro et al. 2010; Jassbi et al. 2018).

To verify the adaptability and the diversity of aggregation methods in this work was performed a comparison of results using seven aggregation operators: Max, Weighted Averaging, Multiplicative and Aditive FIMICA (Yager et al. 1998), Continuous Reinforcement Operator (CRO) (Jassbi et al. 2018), Mean and Weighting Functions, that will be explained ahead. It is important to note that, the chosen operators for comparison, range from the simplest Max operator, to an averaging ones and then a full reinforcement operator, this is, they can be divided in three different types: Simple methods, Average methods and Reinforcement methods (described in Fuzzy Sets Operations, in Section 3.1.1).

3.1.3.1 Simple methods

Max aggregator

Like the name indicates, this aggregation operator uses the maximum values, this is, when aggregating (fusing) data, the best decision is always the one with a higher value. For example, when examining the set of vegetation indices, looking to the membership value of each pixel, with max aggregator, the chosen pixel will be the one with the maximum value, through all indices.

Mean

This aggregation operator calculates the mean values. The data is aggregated following the mean value, i.e., when examining the set of vegetation indices, looking to the membership value of each pixel, the mean operator calculates the mean value for each pixel of all vegetation index, and then assign the output pixel as the mean value obtained before.

3.1.3.2 Averaging/Weight operators

Weighted Averaging

Relative importance (weights) of each criteria enable the penalization of low degrees of criteria performance and the rewarding the high ones when we perform the aggregation average (Ribeiro et al. 2003; Beliakov et al 2007; Calvo et al. 2012). Weighted averaging falls in this concept, this is, is a procedure where each observation in the data set is assigned a weight (importance) before the average operation.

Weighting Functions

A weighting function is a mathematical formalization used to assign weights to each input data, that depend on the satisfaction of the respective criterion These functions were described in step 4 of FIF (section 3.1.2) Reinforcement operators

FIMICA aggregation operator

These aggregation operators are derived from MICA ones (Manyika et al. 1994) and exhibit full reinforcement behavior. They have a value g associated, called identity element, which allow to define levels for good and bad scores, above and bellow g. The smaller the g is, more scores tend to have an aggregation result close to 1.

Being F a FIMICA operator, F has the following properties (Yager et al. 1998):

- Monotonicity - $F(a,b) \leq F(c,d)$ *if* $a < b$ *and* $c < d$.
- Identity - There is an identity element $g \in]0,1]$.

FIMICA operators have two kinds of families, additive and multiplicative (Yager et al.1998), which are depicted in Figure 3.17 and 3.18, respectively, their formulation.

-Additive FIMICA

$$M(\mathbf{x}) = f\left(\sum_{i=1}^{n} (x_i - g)\right)$$

Figure 3.17: Additive FIMICA function (G. Campanella, R.A. Ribeiro / Decision Support Systems 52 (2011) 52–60).

-Multiplicative FIMICA

$$M(\mathbf{x}) = f\left(\prod_{i=1}^{n} \frac{x_i}{g}\right)$$

Figure 3.18: Multiplicative FIMICA function (G. Campanella, R.A. Ribeiro / Decision Support Systems 52 (2011) 52–60).

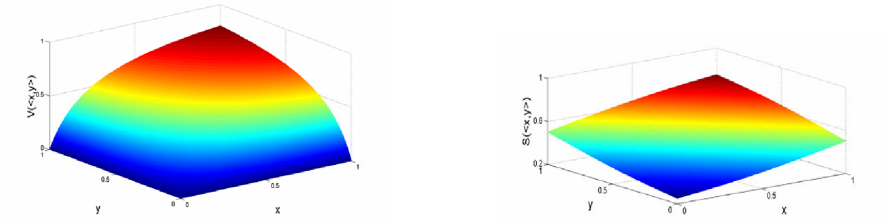


Figure 3.19: Plot of product (left) and additive (right) FIMICA operators, for g= 0.5 (Greco et al. 2010).

An important aspect of FIMICA operators is the choice of an appropriate representative function, f, making it possible, for example, to avoid undesired asymptotic behaviors, controlling the operator behavior (e.g. small changes in the argument resulting in huge differences in the aggregated value) (Yager et al. 1998). In (Ribeiro et al. 2010) were tested several functions for Additive and Multiplicative operators. Examples of those functions are depicted in Figure 3.20, where S function is for Additive Fimica, while V function is for Multiplicative Fimica.

$$S(<x,y>) = 0.5 - \frac{\arctan((x-g)+(y-g))}{\pi} \quad \text{, where } x,y \in [0,1]$$

$$V(<x,y>) = 1 - \frac{1}{1+\frac{x}{g}\times\frac{y}{g}} \quad \text{, where } x,y \in [0,1]$$

Figure 3.20: Functions applied to Fimica operator (Ribeiro et al. 2010).

Continuous Reinforcement Operator

Continuous Reinforcement Operator (CRO) (Jassbi et al. 2018), is also a full-reinforcement operator. CRO is an improved version of a previous reinforcement operator, multiplicative FIMICA. It also has an identity element g ($g \in\ ]0, 1[$), which represents, as aforementioned, the parameter controlling the reinforcement level, upward or downward, and $xi > 0$, for all criteria (Jassbi et al. 2018).

$$CRO(x) = \frac{g^n+1}{\frac{g^n}{\prod x_i}+1}$$

Figure 3.21: CRO aggregation operator (Javad et al 2018).

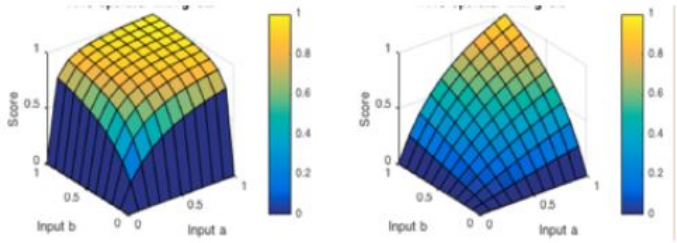


Figure 3.22: CRO behavior with g=0.2 (left) and g=0.8 (right) (Jassbi et al. 2018).

The behavior of CRO using 2 variables and 2 control parameters (g) can be seen in Figure 3.22. The input variables (a and b) values range is between 0 and 1. The aggregated values using CRO, of each example, are represented by their score (vertical axis). If we set the threshold g=0.2, which means high upward reinforcement, a largest set of solutions are acceptable. Conversely, if we set g=0.8 (high downward reinforcement) only the results with very good satisfaction levels in the inputs will be acceptable.

3.2 Summary

Over the years, remote sensing systems are becoming a fully developed field. There are many classifiers and different technologies used in remote sensing problems, producing very promising results, like Artificial Neural Networks, Decision Trees, among others. These improvements lead to the introduction of the concept of data fusion, which consists of merging multiple images to obtain a single component.

As recent studies demonstrated, fuzzy logic technologies applied to remote sensing data fusion are proving to be a useful approach in order to deal with classification studies, such as land cover (Mora et al. 2017), where several approaches were compared.

Having revealing such promising results, for this work, a fuzzy data fusion approach is used for land cover classification, more specifically, for eucalyptus trees identification.

Chapter 4

System Development

This chapter will be divided in two sections. The first one describes how the created system was developed and the information about context and tools used. The second section describes the architecture behind the developed system and how the data preparation is performed. After, it is discussed how the results can be reproduced.

The work is centered in two technologies: Python, which allows handling the input data to achieve the final results; and QGIS, which is a Geographic Information System (GIS) software. Both are briefly described bellow.

4.1 Used Technologies and Tools

4.1.1 QGIS

QGIS is a user-friendly Open Source Geographic Information System (GIS) licensed under the GNU General Public License. It runs in most well-known operative systems and supports numerous vector, raster, and database formats and functionalities. Regarding this work, QGIS was used for the download and preprocessing of the Sentinel 2 imagery, for the calculation of the used vegetation indices and for producing the final input data (tiff rasters) to be used.

This tool was chosen to develop the proposed problem because QGIS is a really good GIS application to access geographic information or even official satellite images. Another interesting quality of this software is that it is open-source and really intuitive and easy to work with. About its features, QGIS accepts files like ESRI shapfiles and mapinfo tab, allowing to map thematically and visualize geotiffs, been the shapefiles easy to edit. Since QGIS is an open-source application, being compatible with all sorts of tools and technologies it is a huge advantage. The most used tools throughout QGIS are automatically ready for usage after the software installation and (another nice feature) - many less used

functionalities or required for specific problems, are offered as plugins and extensions by the community or by the software developers connected to the QGIS development. These plugins (or extensions) allow usage to functionalities like georeferencing to image geocoding, support to visualization and editing OpenStreetMap data, between a large set of other features. The last version came to public in August 16, 2019, with QGIS 3.8.2 "Zanzibar"and also includes one of the most active communities, concerning open-source tools.

Next it will be described the main tools/concepts used regarding QGIS software.

Semi-Automatic Classification Plugin

To perform an analysis and to draw some conclusions about the eucalyptus tree identification process, Sentinel 2 satellite images were needed, as mentioned. The used images are from March 11, 2019 and were extracted directly from QGIS software, using a feature called Semi-Automatic Classification Plugin (SCP). This functionality can be accessed by installing this plugin, inside QGIS. It allows to select the map area that we want to extract from the Sentinel 2 (Landsat, MODIS, etc.) images, drawing a polygon over the QGIS map. After defining the desired area, it is possible to choose between the various days of the year (or other years) available, the percentage of cloud cover present in the images, as well as if we want to preprocess the images or not. After this step, the images (all bands in our case) are downloaded, ready for usage. The images are downloaded as raster data (*tiff* files).

Raster Data

In GIS systems, raster data consist in matrices of discrete cells, representing features on, above or below the earth's surface. These matrices of cells (or pixels) are organized into rows and columns, where each cell contains a value representing information (for example, temperature). Each cell in the raster grid is of the same size, which in QGIS they are always rectangular. Typical raster datasets include remote sensing data, such as aerial photography, or satellite imagery.

The images are geocoded by pixel resolution and the x/y coordinate of a corner pixel of the raster layer. So, this allows QGIS to position the data in the right way in the map canvas. Inside the raster layer, QGIS makes use of georeference information, like *GeoTiff*, for example.

In this work raster data was fundamental, since it is the source type of the downloaded satellite images, where each pixel of those images has a value associated. Hence, it allowed to calculate the vegetation indices using another QGIS tool, *Raster calculator*, which ensures the possibility of performing calculations with different raster images (not only), therefore, producing the final tiff images to be processed.

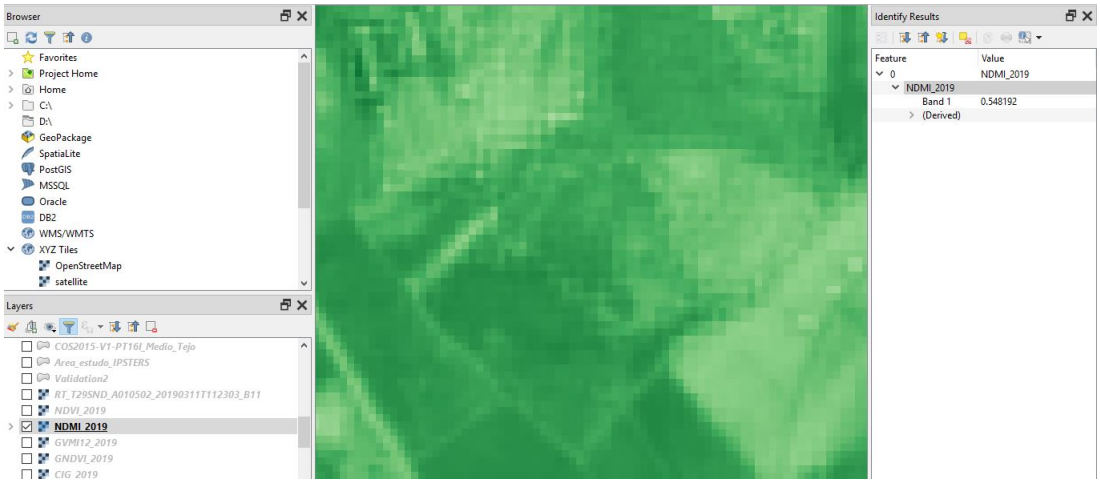


Figure 4.1: Raster layer example (Source: QGIS 3.4.5).

Figure 4.1 depicts a raster representing a vegetation index, previously calculated (NDMI). As can be seen, in the right window, the values of each pixel, regarding raster info, can be accessed. This happens clicking in a pixel of the raster.

Shapefiles

Represents a very popular vector data format of GIS software, mostly developed and maintained by Esri (Environmental Systems Research Institute), which is a supplier of GIS software, web GIS and geodatabase management applications. This kind of data describes vector features, like points, polygons and lines. Each element has attributes associated, such as a name or an ID. Shapefiles have many advantages, like fast drawing speed and edit capacity, since do not have the processing overhead of a topological data structure (ESRI Shapefile Technical Description 1998). Having a large set of characteristics, they can be used in many situations, for instance, in a map it can be draw points shapefiles representing houses, line shapefiles representing roads and polygon shapefiles representing a set of areas.

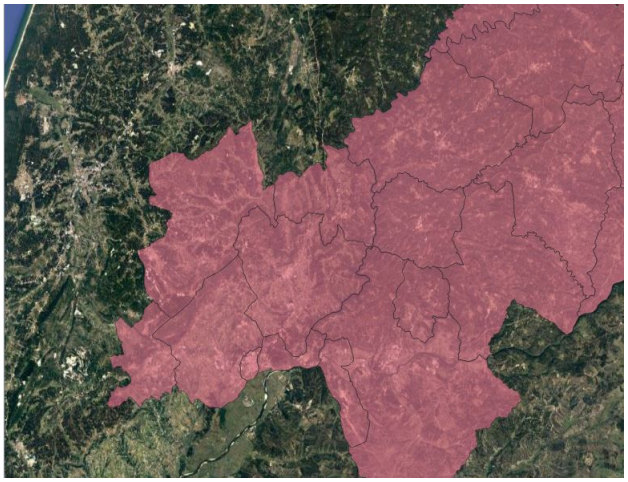


Figure 4.2: Shapefile in map example (Source: QGIS 3.4.5).

Figure 4.2 shows an example of information represented by a shapefile applied to a map, which is represented by the polygons in pink color. This shapefile gathers a set of polygons, delimiting a certain area.

Shapefiles are not represented only by one file, but by a set of files. There are three mandatory ones:

- *.shp* - Represents the shape format, which is the feature geometry.
- *.shx* - Represents the shape index format.
- *.dbf* - Represents the attribute format.

This type of format was fundamental for this work for two main reasons. Firstly, polygons shapefiles were used to draw the areas to be processed and analysed (for classification). It was necessary to selected parts of the raster images, where the "cut" (selected area) represents the area to be processed. Polygon shapefiles are also very useful for this situation, since QGIS enables the possibility of cutting a raster file based on a polygon shapefile, i.e. the resulting raster, after the cut, will be only the area of the polygon. Secondly, for the validation step, were used polygon shapefiles, provided by COS, representing areas that are in fact eucalyptus, allowing the achievement of the ground truth result.

4.1.2 Python

Python is a well-known and widely used language and consists in an interpreted, high-level, object-oriented, with a wide range of purpose programming fields, from statistical computing to image processing, which will be used in this work. It has available a large set of functionalities provided by various packages. The main packages that were used in this approach were: *matplotlib.pyplot*, *numpy*, *seaborn*, *skfuzzy*, *math*, *PIL* and *osgeo*. From *osgeo* package were used two "sub-packages", *gdal* and *gdal_array*.

osgeo

Represents the fundamental package for the development of this work, withits two "sub-packages", used to read and transform the input data (*tiff* files) into comparable and numerical data.

osgeo offers two libraries: GDAL and OGR. GDAL one was used for manipulating geospatial raster data (tiff files), while OGR (OpenGIS Simple Features Reference Implementation) is used for manipulating geospatial vector data. GDAL acronym in *gdal* and *gdal_array* sub-packages stands for Geospatial Data Abstraction Library.

4.2 Work Approach

1. The set of Sentinel 2 satellite images for our study area are from March 11, 2019 and were acquired from online archives (through QGIS software, as mentioned in Section 4.1.1).

2. Having Sentinel 2 imagery, we had access to information about all thirteen different multispectral bands. With that information, all six vegetation indices chosen were calculated.

3. Based on our ground truth information (COS *shapefiles*, 2015), the study area was divided into two parts: the first part does not include only eucalyptus trees, to better access the classification results. The second area , is relatively small to enable analyzing in higher detail and draw more thorough conclusions. This second sub-area was drawn/selected using a polygon *shapefile*, by hand.

4. Each vegetation index was calculated from the study area *shapefile* in order to export them to be treated as input data.

5. The generated images, regarding vegetation indices information, were imported into Python environment and subsequently processed, applying the two major steps of the data fusion approach: Normalization and Aggregation (described in detail in Chapter 4).

6. The final result is presented in image format (*png*) by matching the aggregated result with the map itself, and this process required to georeference the output image.

4.3 Used System

As mentioned in chapter before, the development of this work was based on two main technologies: QGIS, for data preparation; and Python, for data analysis and classification (image processing). For QGIS it was used QGIS 3.4.5 "Madeira", for Python the environment - where the scripts were written - was Spyder (The Scientific Python Development Environment), using Python version 3.6.

4.3.1 System organization

The organization of the developed work is as follows. The data preparation in QGIS was done independently of the Python script. There is only one folder, which contains all the needed files: the QGIS generated data (images all of the vegetation indices used, regarding the study area); and the Python script that contains the code that handle all of the classification procedure.

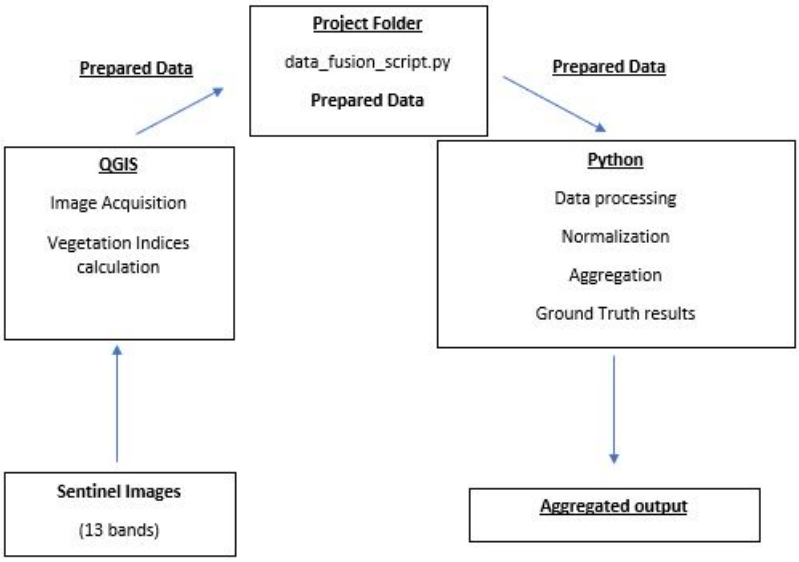


Figure 4.3: System organization.

The satellite images are downloaded and imported to the QGIS software and then are prepared, like Figure 4.3 suggests.After this process, the prepared images (as raster data) are exported and saved in the project folder, which contains the script used to run and execute the proposed work, *data_fusion_script.py*. Then, the script imports the images that were saved previously, corresponding to input data.

About the developed script (***data_fusion_script.py***), it loads the images as raw data, extracts the values of each pixel for each vegetation index and saves them in an array. At this point, there is the need do some preparation of the input data. Then, for each of the vegetation indices arrays, it is calculated their membership functions, process called Normalization, (described in detail in Section 5.1.2) and saved in an array. The final process is to aggregate all of the obtained membership functions, guided by the aggregation operators (Aggregation process), as described in Section 5.1.3.

4.4 Data Preparation

There is a need for data preparation before the actual classification is done, as mentioned before. Thus, this task can be divided into two parts: Image Acquisition and Vegetation Indices Calculation; and Image Cleaning.

4.4.1 Image Acquisition and Vegetation Indices Calculation

This stage is completed using QGIS and at this point, the aim is to obtain the Sentinel 2 satellite images. As mentioned before we use the SCP plugin, which preprocess and download Sentinel 2 images in raster layers format (*tiff* files). Firstly, we specified the

area to be downloaded, drawing a polygon with SCP tool, to obtain the right coordinates. The preprocessing consists in an operation correcting radiometric, geometric and atmospheric aspects to improve the interpretation of image components, quantitatively and qualitatively (Sowmya et al. 2017). Second we chose options to download different images including the chosen area, as well as other relevant information, such as cloud cover percentage, date, etc. The last step is to choose the intended image and confirm the download. The final result will be the download data of the thirteen bands of the chosen Sentinel 2 image.

The second step corresponds to the Vegetation Indices calculation with data from the thirteen multispectral bands (Section 2.1.3). For this, it is used another QGIS tool named *Raster calculator*, which allows mathematical operations with the different raster layers that are open in QGIS environment.

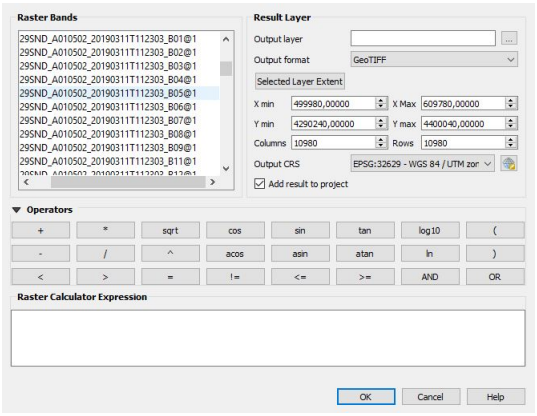


Figure 4.4: Raster calculator (Source: QGIS 3.4.5).

As can bee seen in figure 4.4, *Raster calculator* is very intuitive. Taking in account the index formula to be calculated, we build that formula by selecting the required bands and use the available QGIS mathematical operators. Finally we just need to select the output folder and the name of the file.

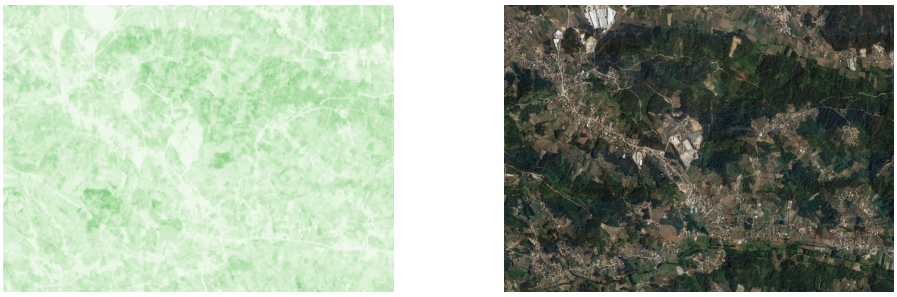


Figure 4.5: CIG index (left) and satellite image (right) regarding the same area (Source: QGIS 3.4.5, 2019) (More details about other used indices on Section 5.1.1).

4.4.2 Image Cleaning

This is the last step regarding data preparation. When the raster images are imported to Python, they include some "noise"values, which instead of representing a pixel value of the index itself they are non-numerical data these values are defined as "NoneType".

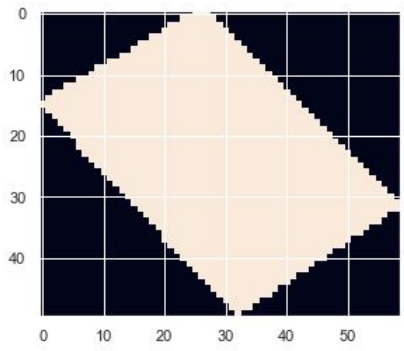


Figure 4.6: Difference between numerical data (pink pixels) and non-numerical (black pixels) (Source: Python (Spyder) 3.6, 2019).

Figure 4.6 depicts an example of the plot for the NDVI vegetation index. As can be seen, there are a large amount of “NoneType” pixels (black ones) and so they need to be removed in order to get all data comparable and numerical. “NoneType”signifies that the variable (pixel in this case) does not hold any value. This is done by identifying these values (supported by GDAL functions) and then construct an array only with numerical data. To note that non-numerical pixels positions are saved to reconstruct the image, maintaining the same size.

Chapter 5

Case Study

In this chapter, the proposed data fusion approach is applied to the case study and the obtained results are discussed. It also includes the validation step to assess the suitability of the approach.

5.1 Data Fusion Approach for Eucalyptus Identification

Figure 5.1 depicts the four steps of our proposed data fusion approach for Eucalyptus trees identification. Notice that step 1 includes the retrieval and part of the preparation of data as explained in Chapter 4. Hence, here step 1 will focus only on the rational for choosing the criteria (vegetation indices) and the bands used for performing the respective calculations.

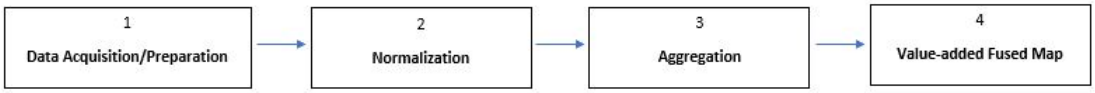


Figure 5.1: Steps of the data fusion approach.

The chosen area for our study includes an eucalyptus forest, as well as other types of land, roads inside the forest, pine forests etc. We believe this area also demonstrates how the vegetation indices behave in terms of differentiating different types of land and what influence each one has in the identification of eucalyptus. The area is located near the locality of Olival that belongs to the city of Ourém, in Santarém District (39.697485, -8.588190).

Figure 5.2: Study area.

Figure 5.2 depicts the study area (inside red polygon) used to present the procedures of the data fusion approach. As can be observed in Figure 5.3, this area represents a "mixed"area, with eucalyptus trees represented in the red spots (COS 2015), non-eucalyptus and a cultivation field (non highlighted zones). In the next sub-section we discuss details of the chosen vegetation indexes and the rational for selecting them.

Figure 5.3: COS map applied to the study area.

5.1.1 Vegetation Data

As mentioned before, for this study (eucalyptus trees), we selected 6 vegetation indices for our data fusion approach: (i) Normalized Difference Vegetation Index (NDVI); (ii) Green Chlorophyll Index (GCI); (iii) Green Normalized Difference Vegetation Index (GNDVI); (iv) Global Vegetation Moisture Index (GVMI); (v) Normalized Difference Moisture Index (NDMI); (vi) Soil Composition Index (SCI). We believe these indices are enough to differentiate Eucalyptus forests, since they match most of the tree specific characteristics. Further, we also used Band 11 as a seven criterion for the data fusion approach. Next, we present the details about the 7 criterion for the data fusion approach.

As mentioned in Chapter 3, for the calculation of the vegetation indices was used *Raster Calculator*, a QGIS plugin. As will be seen, some indices use bands with different resolutions. A big advantage of *Raster Calculator* is that, when we calculate the indices with different band resolutions, it converts the output automatically in one single resolution and, in our case produced 10x10 pixel resolution images for the indices.

- NDVI
  The Normalized Difference Vegetation Index (NDVI), introduced in Section 2.1.3, measures vegetation health, i.e. green vegetation (Rouse et al. 1973). The index is calculated measuring the difference between near-infrared (where vegetation strongly reflects) and red light (where vegetation absorbs) and the bands used were Band 8 (NIR) resolution and Band 4 (Red), with 10 meters resolution.

$$NDVI = \frac{(NIR - Red)}{(NIR + Red)}$$

  The values range is -1 to +1. For example, when we have negative values, it is highly possible that it is water. Conversely, if we have a NDVI value close to +1, there is a high chance that it refers to dense green leaves. Considering that eucalyptus forests, when they reach the adult age (and are healthy), display the behavior of very dense forests with dense green leaves, NDVI helps to identify them when values are closer to +1. Figure 5.4 depicts the response of the NDVI index, applied to two different kinds of land (inside red polygons). The left images are satellite images and their corresponding NDVI values are on the right. The top images correspond to eucalyptus trees (values tending to +1), while bottom images represent water (values near -1) and agriculture fields.

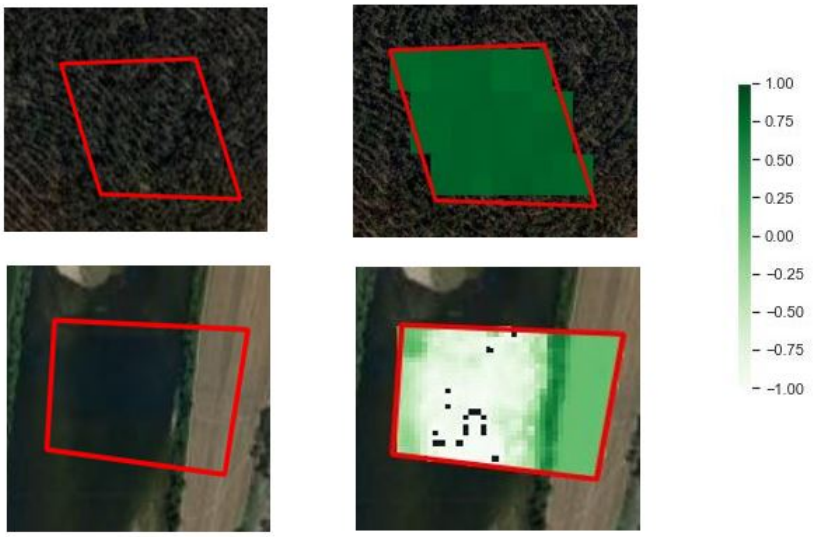


Figure 5.4: Differences in NDVI index, changing the type of cover.(Source: QGIS 3.4.5)

- CIG
  Green Chlorophyll Index estimates the chlorophyll content in leaves using the ratio of reflectivity in the near-infrared (NIR) and green bands.

$$CIG = \frac{NIR}{GREEN} - 1$$

  Studies show that eucalyptus display a relevant value of chlorophyll content (Coops et al. 2003), therefore, a high value for this index indicates it is possible it is an eucalyptus. Again, note that the calculations are done per pixel. Here it was used Band 9 (NIR), with 60 meters resolution, and band 3 with 10 meters resolution. The best values for identifying eucalyptus trees lie between 9 and 11.

- GNDVI
  Green Normalized Difference Vegetation Index is very similar to NDVI but with some differences. Firstly, it measures the green spectrum from 540 to 570 nm instead of the red spectrum. Secondly, this index is more sensitive to chlorophyll content than NDVI (Gitelson et al. 1996). For this index, the values that we found being the best for eucalyptus trees are between 0.76 and 0.86. The bands used for calculating this index were Band 9 (NIR) with 60 meters resolution, and Band 3 (Green) with 10 meters resolution.

$$GNDVI = \frac{(NIR - Green)}{(NIR + Green)}$$

- GVMI
  Global Vegetation Moisture Index provides information about the vegetation water content from an area (Ceccato et al. 2002). Since eucalyptus are trees that have very high water consumption levels (Liu et al. 2010), as well as leaf water content (Bisun Datt, 1999), this index is very useful to identify them. Band 9 (NIR), with 60 meters resolution and Band 12 (SWIR), with 20 meters resolution were used for calculation of this index. The acceptable range of values for eucalyptus identification are [0.39 ,0.71].

$$GVMI = \frac{(NIR + 0.1) - (SWIR + 0.02)}{(NIR + 0.1) + (SWIR + 0.02)}$$

- NDMI

  Normalized Density Moisture Index describes the crop's water stress level and is calculated with a ratio between the difference and the sum of the refracted radiations in the NIR and SWIR.

$$NDMI = \frac{NIR - SWIR}{NIR + SWIR}$$

  NDMI recognize the areas of vegetation with water stress problems. Its values vary between -1 and 1 (like NDVI), and each value corresponds to a different agronomic situation, for example, value -1, or near, represents bare soil, while 1 (or near) represents total canopy cover with no water stress/waterlogging (https://www.agricolus.com/en/indici-vegetazione-ndvi-ndmi-istruzioni-luso/). The best values to identify Eucalyptus are in the range [0.2 - 0.5] and Band 8 (NIR), with 10 meters resolution and Band 11 (SWIR), with 20 meters resolution were used for the index calculation.

- SCI

  Soil Composition Index is used to spot the differences between what is soil and what is vegetation. In the proposed work, this index was very useful to remove data that is not vegetation (roads, fields, etc.). Since we use membership functions, the lower a pixel is classified as vegetation i.e. the closer to 0 in the membership function. Therefore, when this index enters in the aggregation step, pixels that are 0 (or near) will have no importance, reinforcing negatively the non-vegetation pixels, which is the wanted output. The best values are between -0.35 and - 0.25. The used bands were Band 11 (SWIR) and Band 8 (NIR), with resolution of 20 meters and 10 meters, respectively.

$$SCI = \frac{SWIR - NIR}{SWIR + NIR}$$

- B11

  This last chosen criterion is the only one that is not a vegetation index. It refers to source data, retrieved directly from the spectral Band 11 of Sentinel 2 satellite. Even though it has limited cloud penetration, it is quite useful for measuring the moisture content of vegetation and it provides good contrast between different types of vegetation - a very important measure for eucalyptus identification. Its central wavelength is found at 1610 nm. (https://www.sentinel-hub.com/eoproducts/band-b11)

Above it is represented a satellite image of a random area, with different land types and forests (left figure) and its representation with Band 11 (right picture). As can be seen, Band 11 value variation depends highly on the type of land, presenting good contrast between different types of vegetation. Band 11 has 20 meters resolution and the best values found were between 0.08 and 0.12.

5.1.2 Step 2 - Normalization

This step corresponds to the transformation of the input data domain into normalized numerical and comparable data, i.e., taking in account all the vegetation indices chosen, the normalization process transforms each image corresponding to each index into the [0,1] domain, using a membership function (Section 3.1.1). Hence, each pixel of the original image, is guided by a membership function that assigns value 1 (higher) to the best classified pixels and 0 (lowest) to the worst pixels.

Since there are 6 different vegetation indices (input data) and 1 band, the normalization process was executed for each one, resulting in 7 membership functions of normalized data. To normalize each index used, we used three types of function, Sigmoid, Gaussian and Trapezoidal, depending on how well they fitted the data from the images used. A summary of the membership functions used and their respective parameters for each vegetation index is depicted in Table 5.1.

This normalization step was based on the specific area selected (see Figure 4.2). The choice of membership functions topologies was based on the analysis of the mentioned area, by retrieving, for each index, the values of the pixels that were eucalyptus trees and then their variance interval (domain) was built.

Variable	Range	Method	Parameters
NDVI	[0.4877,0.9025]	Sigmoid	Offset = 0.75 Width = 30
CIG	[2.3402,12.7055]	Trapezoidal	a (lower limit) = 7 b (upper limit) = 9 c (upper support limit) = 11 d (lower support limit) = 14
GNDVI	[0.53919,0.86399]	Sigmoid	Offset = 0.76 Width = 30
GVMI	[0.3577,0.7146]	Sigmoid	Offset = 0.39 Width = 30
NDMI	[-0.0233,0.4954]	Sigmoid	Offset = 0.2 Width = 30
SCI	[-0.4954,0.0233]	Gaussian	Mean = - 0.28 Sigma = 0.25
B11	[0.0666,0.2031]	Gaussian	Mean = 0.1 Sigma = 0.05

Table 5.1: Information regarding the normalization step.

In the cases where sigmoid functions were applied (NDVI, GNDVI, GVMI, NDMI), the membership value would be always increasing, i.e., the value was set as offset of the sigmoid function. For example, Figure 5.5 shows the membership for NDMI index.

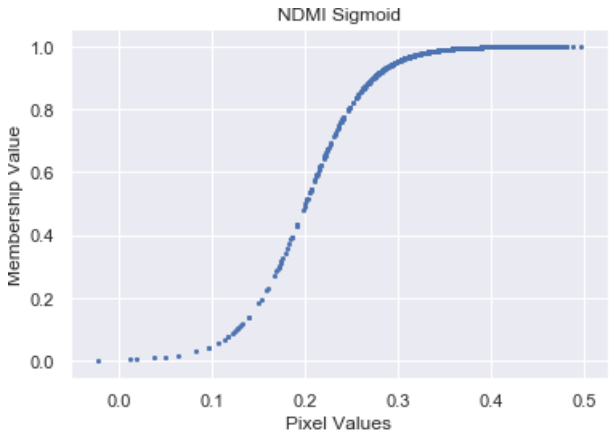


Figure 5.5: Sigmoid function for NDMI index.

For the Gaussian membership functions the process was a little bit different (Indices SCI and B11) . For example, in B11 index, the pixels classified as "good"membership values fall within a variation interval (0.08 to 0.12) and then higher values (like 0.15) refer to areas that we not want to be assigned as eucalyptus. So, the Gaussian membership function is appropriate because the membership value increases until its center values (mean) and then starts to decrease, which was exactly what was pretended for SCI and B11.

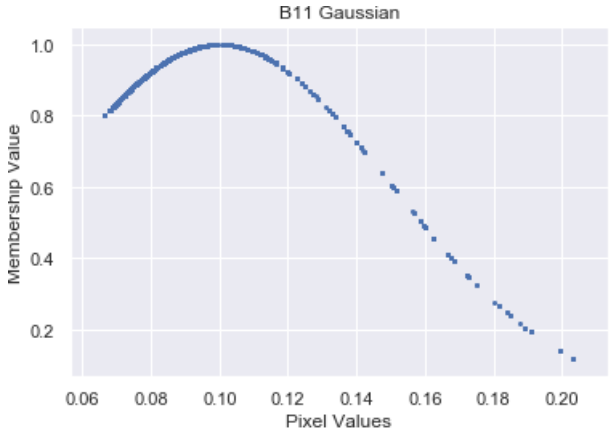


Figure 5.6: Gaussian function for B11.

Figure 5.7 depicts the behavior of the trapezoidal function selected for CIG. The best values are the pixels with values between 9 and 11.

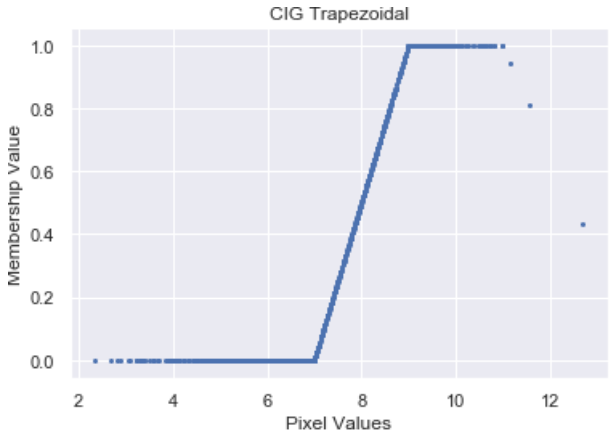


Figure 5.7: Trapezoidal function for CIG index.

The normalization process was applied to all of the chosen indices and Figure 5.8 illustrates the complete process of normalization (step 2 of the data fusion approach) with an image of the original image and then, its respective membership function and the final result, the normalized image.

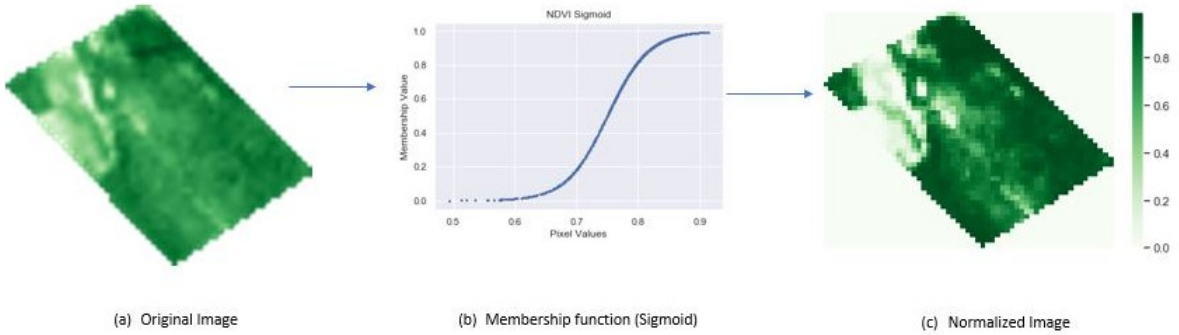
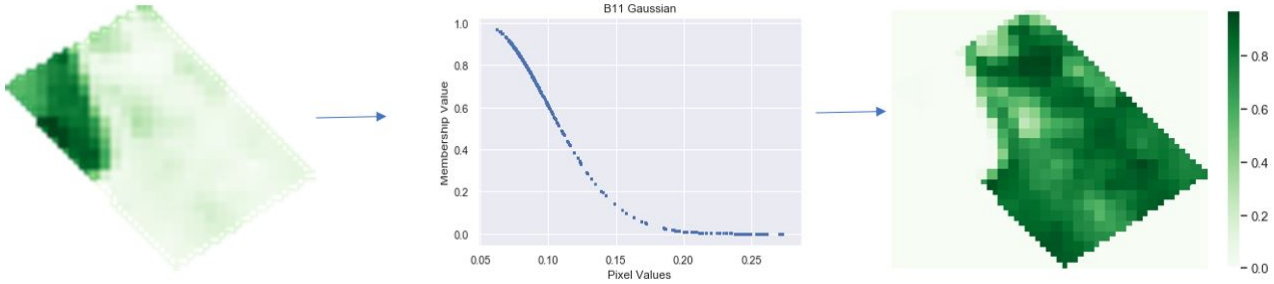


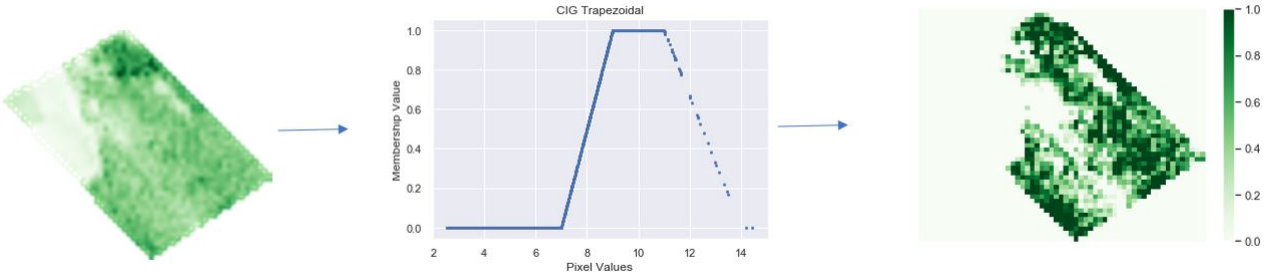
Figure 5.8: NDVI index normalization process (Source: Python 3.6).

In Figure 5.9 is depicted the results of the normalization process applied to the rest of the indices, where the left represent the original images of the index, the middle ones are their corresponding membership functions and the right ones the normalized images.

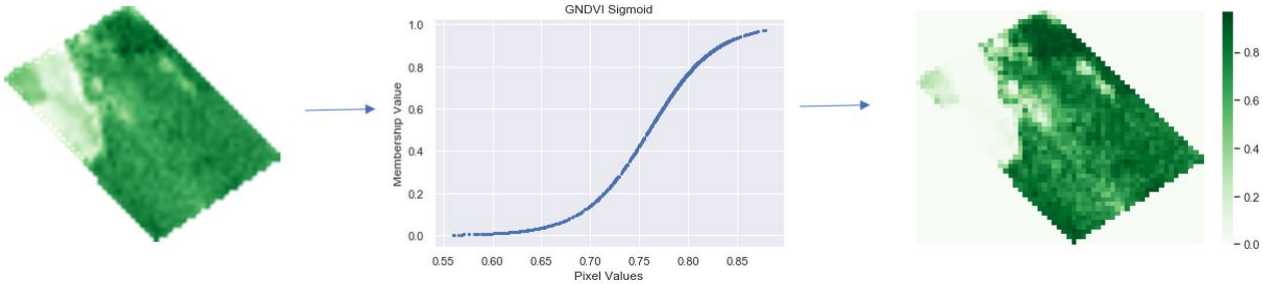
Band 11
(Gaussian function)



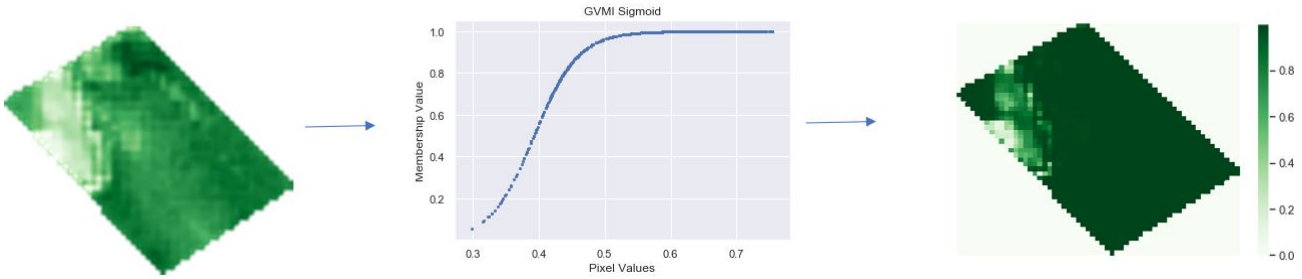
CIG
(Trapezoidal function)



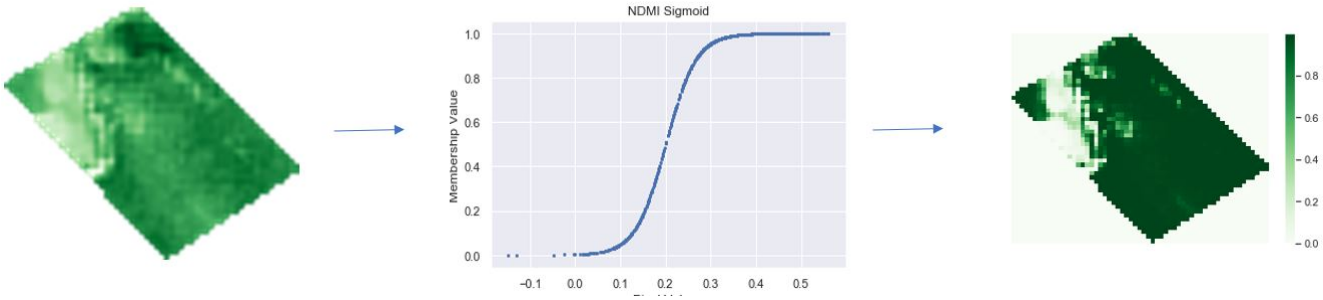
GNDVI
(Sigmoid function)



GVMI
(Sigmoid function)



NDMI
(Sigmoid function)



SCI
(Gaussian function)

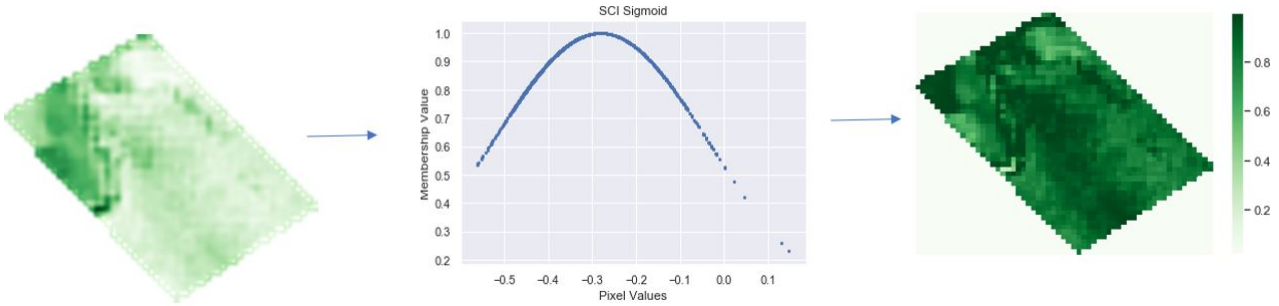


Figure 5.9: Normalization process applied to all used vegetation indices.

As can be seen, applying membership functions to the indices original images changes significantly its properties, regarding its original values. Assigning membership values to each pixel gives a very useful way to aggregate all the data (will be described in the next section), since makes it possible to know the information of which pixels are better for each index.

5.1.3 Step 3 - Aggregation/Fusion

This step consists in the aggregation/fusion of the obtained normalized data to create a single added-value map/image with highlighted Eucalyptus trees. As mention in section 3.1.3 we tested and compared the results with 7 different aggregation operators from the three classes (algebraic, average and reinforcement): (i) Max; (ii) Mean; (iii) Weighted Averaging; (iv) Weighting Functions; (v) Continuous Reinforcement Operator (CRO); (vi) Multiplicative FIMICA; (vii) Additive FIMICA.

With the objective of finding the best aggregation operator, we gathered ground truth data regarding the tested area. It is rather important to highlight that the ground truth is made manually by DGT and here we discuss an automated process with our data fusion approach.

It is important to have ground-truth information to confirm, in the tested area, without doubts, which pixels corresponded in fact to eucalyptus trees. We obtained this information from COS *shapefiles* after converting this *shapefile* to the same format as the output result, i.e. same resolution. With the obtained ground truth data we manage to validate, for each aggregation operator, the percentage of correct classifications.

Figure 5.10: Study area satellite image (left), inside the red polygon, and the ground truth data image (right), where the red area represents eucalyptus trees, as *shapefile* (Source: QGIS 3.4.5, 2019).

Ground Truth Tranformation

As mentioned before, we transformed the *shapefile* data regarding the ground truth to obtain compatible data with the our output results. First we transformed the *shapefile* information into a georeferenced image, with respect to the study area. Second, the generated georeferenced image was transformed into a raster layer (tiff format). Third, we set the resulting raster layer resolution equal to the output one (10x10 pixels), to make both comparable.

Figure 5.11: Original *shapefile* image (left) and its transformation result (right) (Source: QGIS 3.4.5, 2019).

In addition, since we are dealing with fuzzy data, we defined a threshold for our solution from which all solutions above number/count as being classified. With this we obtained the counting of classified as Eucalyptus and the number/count of not classified as Eucalyptus. The chosen threshold value was 0.3 because, by observation, it seems a good transition point, regarding the used aggregation operators.

5.1.4 Step 4 - Value-added maps

Here we present graphically the results obtained for each of the seven used aggregation operators. To enhance visibility of classification results, the left image is always the ground truth and the right ones are the results for each specific operator tested:

Max

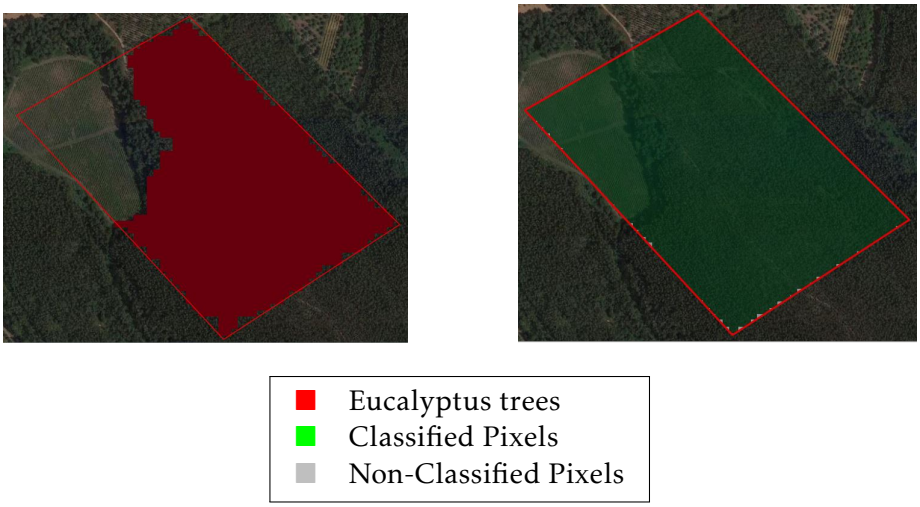


Figure 5.12: Max operator results on image (right image) (Source: QGIS 3.4.5, 2019).

As can be seen in Figure 5.12, the max operator classified almost all pixels.

Mean

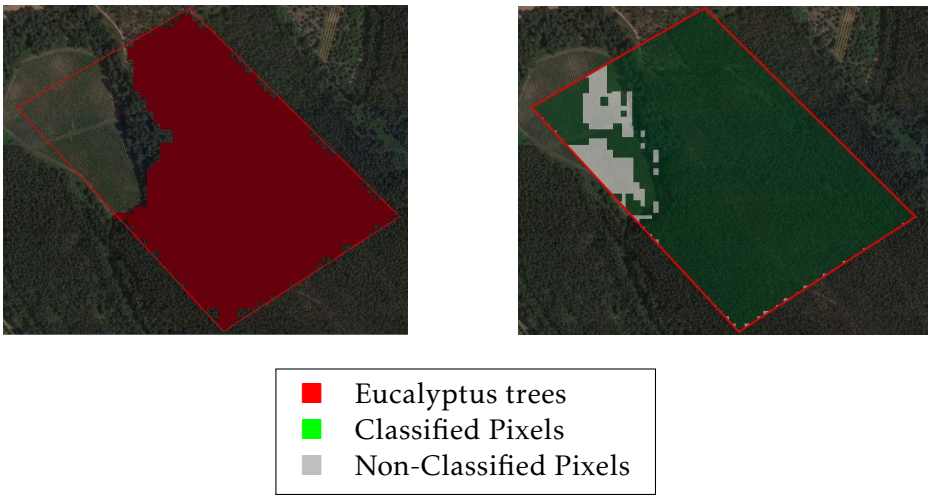


Figure 5.13: Mean operator results on image (right image) (Source: QGIS 3.4.5, 2019).

Weighted Averaging

As mentioned in Section 3.1.3, in weighted averaging we need to define weights for each criteria involved.. With a preliminary visual analysis of the normalized images, and taking in account which pixels are eucalyptus trees (COS), the weights assigned are: 0.3 to NDMI index, 0.2 to CIG index and 0.1 to the rest. To note that the sum of the weights must be equal to 1 (0.3 x 1 index + 0.2 x 1 index + 0.1 x 5 indices = 1).

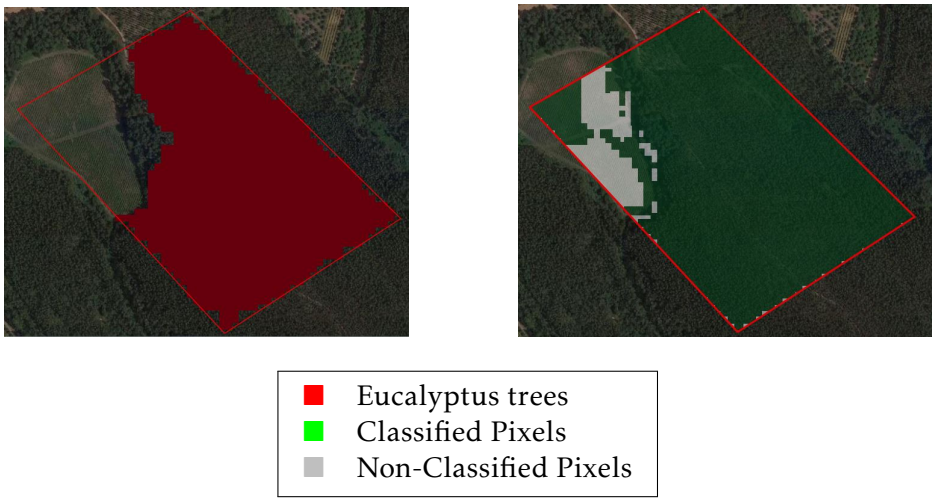
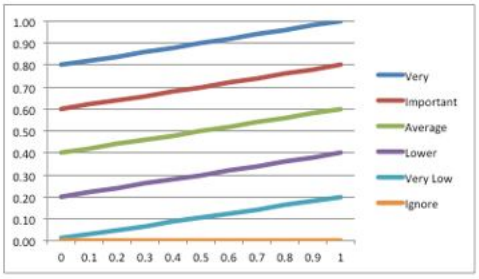


Figure 5.14: Weighted Averaging operator results on image (right image) (Source: QGIS 3.4.5, 2019).

Weighting Function operator

$$L(fu_{ij}) = \alpha \frac{1+\beta fu_{ij}}{1+\beta}, \text{where } 0 \leq \alpha \leq 1 \text{ and } 0 \leq \beta \leq 1$$

The used weighting functions were borrowed from (Ribeiro et al. 2014), and Figure 5.15 displays their morphology. . More details about the weighting functions operator can be seen in Section 3.1.2 (Step 3).



Relative importance intervals	Upper Limit	Lower Limit
Very Important	1	0.8
Important	0.8	0.6
Avg importance	0.6	0.4
Low importance	0.4	0.2
Very Low importance	0.2	0.0
Ignored	0	0

Figure 5.15: Weighting functions used to assign weights to each criterium (Ribeiro et al. 2014).

Regarding our problem, were assigned very important relative importance to NDMI index, important to CIG and low importance to the rest of the input data, according to Figure 5.15.

The process goes as follows. For each pixel of each criteria its satisfaction value (from x-axis) is weighted with its respective weight from y-axis. After, using the weighting function of Figure 5.15 we obtain the results using the assigned weights, as shown below,

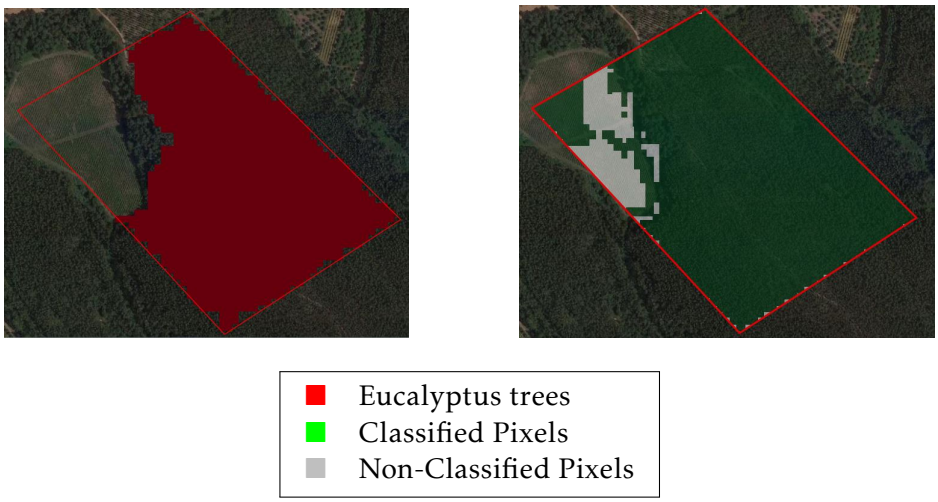


Figure 5.16: Weighting Function operator results on image (right image) (Source: QGIS 3.4.5, 2019).

Reinforcement operators (CRO, FIMICA multiplicative and additive)

For the reinforcement operators, we set the *g* parameter (Section 3.1.3),to a common value for all three used reinforcement operators. This parameter controls the reinforcement level by penalizing scores values below a certain threshold, i.e. the value of neutral element (parameter g) and rewarding values above. We chose a low value (0.1) because we want to accept a sufficiently large number of potential candidates (hit rate), i.e if the reinforcement level is high, fewer pixels are defined as "acceptable". In the futures other g levels will be tested. Figures 5.17, 5.19 and 5.21 depict the results of the three reinforcement operators:

g	Multiplicative FIMICA		Additive FIMICA		CRO	
	Hit Rate	Miss Rate	Hit Rate	Miss Rate	Hit Rate	Miss Rate
0,1	0.829	0.158	0.989	0.641	0.785	0.155
0,2	0.829	0.158	0.984	0.538	0.752	0.155
0,3	0.829	0.158	0.966	0.452	0.715	0.152
0,4	0.829	0.158	0.942	0.333	0.694	0.147
0,5	0.826	0.158	0.869	0.194	0.672	0.144
0,6	0.820	0.158	0.714	0.147	0.659	0.144
0,7	0.790	0.155	0.402	0.102	0.642	0.144
0,8	0.719	0.152	0.002	0.0	0.631	0.144
0,9	0.568	0.138	0.0	0.0	0.620	0.144

Table 5.2: Hit and miss rates varying *g* parameter.

Table 5.2 presents the rates regarding each of the used reinforcement operators, depending of *g*. As can be seen, for all of the operators, the value 0.1 was the most acceptable in most cases. Hit rate measures the number of pixels classified correctly, while miss rate measures the rate of misclassified pixels. More on this will be explained in Section 5.2.

Continuous Reinforcement Operator

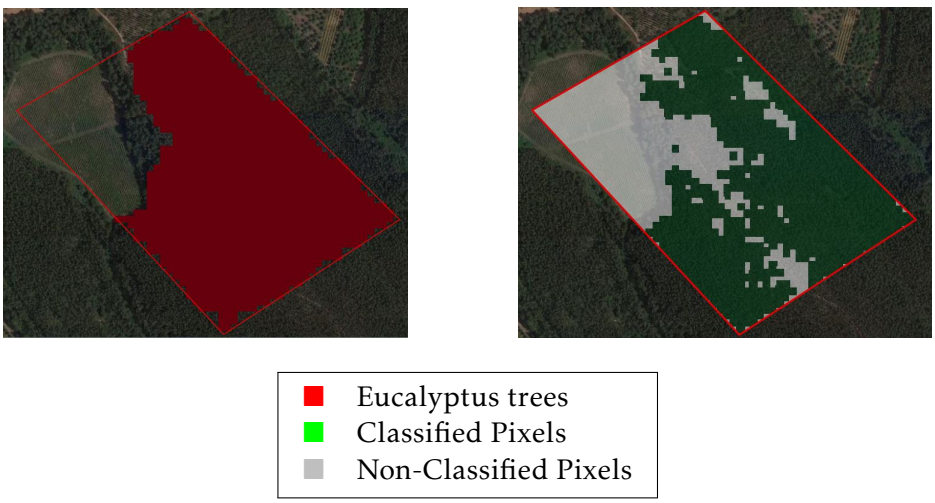


Figure 5.17: CRO operator results on image (right image) (Source: QGIS 3.4.5, 2019).

Figure 5.18: CRO operator plot representing hit and miss values varying *g* parameter (Source: Python 3.6).

Additive FIMICA

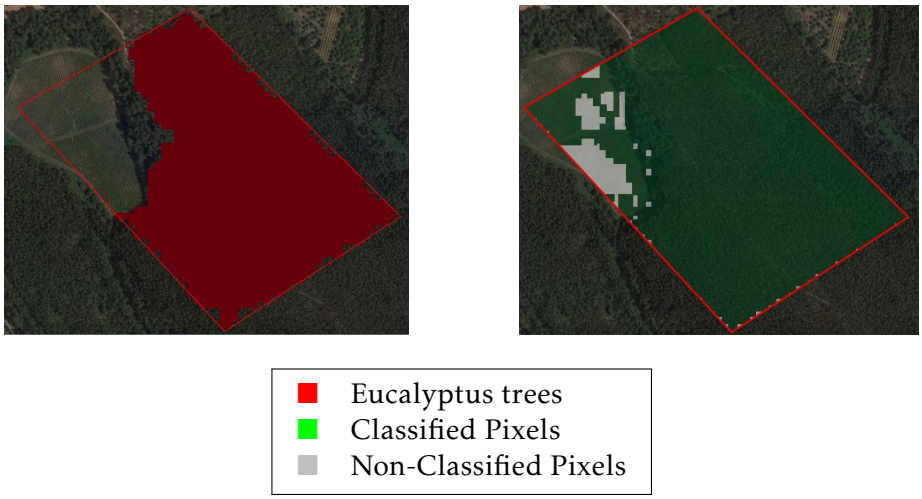


Figure 5.19: Additive Fimica results on image (right image) (Source: QGIS 3.4.5, 2019).

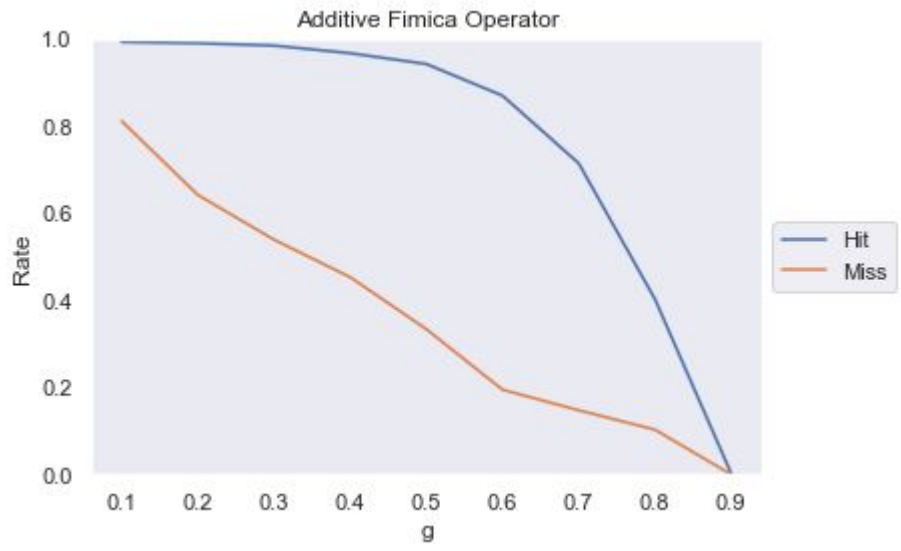


Figure 5.20: Additive Fimica plot representing hit and miss values varying g parameter (Source: Python 3.6).

Multiplicative FIMICA

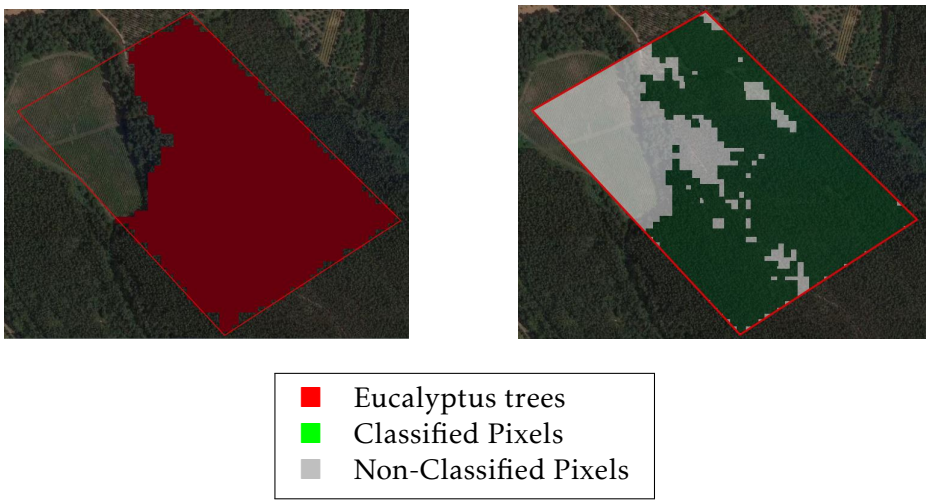


Figure 5.21: Multiplicative Fimica operator results on image (right image) (Source: QGIS 3.4.5, 2019).

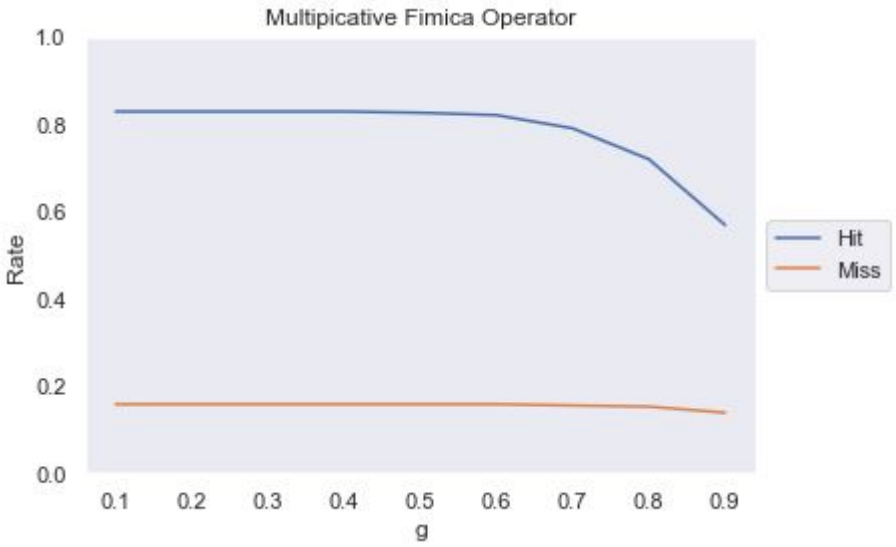


Figure 5.22: Multiplicative Fimica plot representing hit and miss values varying g parameter (Source: Python 3.6).

5.2 Results Discussion

In the previous section we presented graphically the set of results obtained (step 4 of the data fusion approach), through the application of all seven operators and now compare the results obtained with each operator to select which aggregation operator is more suitable for this work. To perform the comparison we use the ground-truth data provided by COS-DGT (the ground-truth is done manually by DGT resulting in the official COS maps).

In order to select which is the best operator we use the counting of Eucalyptus from the ground-truth and compare with the counting provided by each operator. Notice, that, for each aggregation operator, we counted how many pixels matched the ground truth data, resulting on a percentage of hit and missed pixels. The matching performed is as follows:

- **Hit rate** - This measures the number of pixels classified correctly, i.e that matched the ground truth data.
- **Miss rate** - This measures the rate of misclassified pixels, i.e pixels that do not represents eucalyptus on the ground-truth data.

Operator	Number of hits	Number of misses	Miss rate	Hit rate
Max	1081	360	**1**	**0.99**
Mean	1077	266	**0.73**	**0.99**
Weighted Averaging	1077	244	**0.67**	**0.98**
Weighting Function	1077	241	**0.66**	**0.98**
CRO	855	56	**0.15**	**0.78**
Additive Fimica	1077	231	**0.64**	**0.98**
Multiplicative Fimica	902	57	**0.15**	**0.82**

Total eucalyptus Pixels	Total non-eucalyptus trees
1088	360

Table 5.3: Obtained results for each operator.

As can be seen in Table 5.3, hit rates are very high, which suggests that most operators correctly classified the eucalyptus trees. However, there is a catch, since we are dealing with a classification problem it is imperative to minimize, as much as possible, the misclassified pixels.

Max operator got the highest hit rate (as well as mean operator), but also has the highest total miss rate, i.e. it classified almost all of the study area, which is a very bad

result. Mean operator, despite presenting a lower miss rate than max, it is still very high for the intended goal.

Regarding averaging operators, both of them (weighted averaging and weighting function) produced very similar results, along with Additive Fimica. All three have very good hit rates but still very high missing rates, hence, resulting in a poor final result.

CRO and Multiplicative Fimica were the ones with the best results, having almost the same miss rate (56 and 57 misses, respectively), which compared with the rest of the operators, it was a very good improvement. Regarding the number of hits, Multiplicative Fimica had almost more 60 well classified pixels, which means an improvement of 4% in relation to CRO. In this case, the decision between both of the operators points to FIMICA multiplicative as the best operator.

In classification problems, at the moment of decision, it is very important to balance these two factors (hit and miss), because we need to minimize misclassification as much as possible, while maximizing the number of hits (CRO and Multiplicative Fimica have a balanced result).

Taking this into account, the final decision is to choose Multiplicative Fimica as the best aggregation operator for the data fusion approach. Notice that in the studied image, the trees that are not eucalyptus, according to COS (green circle), were not classified as eucalyptus by our algorithm, as well as the cultivation field, as can be seen in Figure 5.23.

Figure 5.23: Information regarding non-eucalyptus forest.

Since we are working with image reflectance information our approach revealed some limitations regarding, for example, the existence of roads between forests or when the eucalyptus forest is very young. Both factors cause reflectance values to change and create some confusion with the soil reflectance.

From the comparison analysis above, on the study area, the operator selected is FIM-ICA multiplicative, however, a more thorough analysis is needed on other areas to ensure the decision is robust and correct. In the next section we performed other types of validation to assess the suitability and versatility of our data fusion approach.

5.2.1 Validation

This section presents the validation of the chosen operator (FIMICA multiplicative) , using five other areas. The chosen areas had in mind different coverage of other representative cases. Exactly the same procedures, as in the previous section (hit/miss rates), are used to assess the robustness of the data fusion approach. Notice that the COS images correspond to the ground-truth provided by DGT.

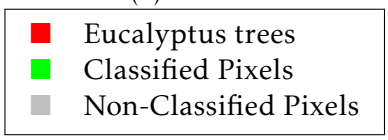
Scenario 1- Only eucalyptus trees

The chosen area for Scenario 1 is an area near the A1 highway, near the village of Vila Nova de São Pedro, belonging to the municipality of Azambuja (39.201693, - 8.809251).

(a) Satellite Image

(b) Eucalyptus area (COS)

(c) Result

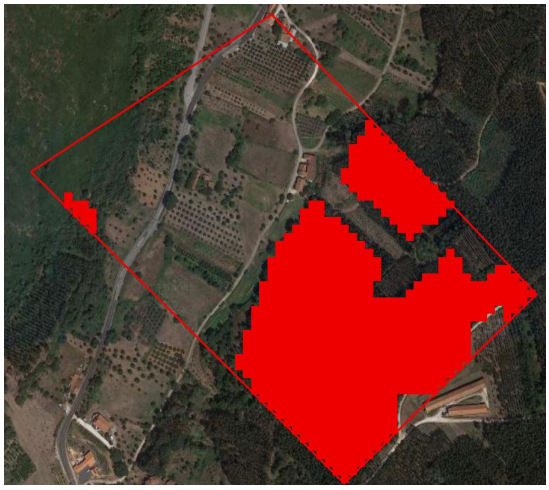


In this case, area only with Eucalyptus trees, the results are very positive. There are some pixels in the top right position that were not classified because there is a large opening on the forest (soil very visible), which may cause confusion between forest and soil.

Hit Rate: 98%
Miss Rate: 0% (since all pixels correspond to eucalyptus)

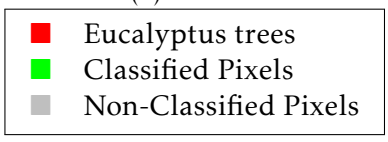
Scenario 2 - Eucalyptus forest with roads, soil and some vegetation

(a) Satellite Image



(b) Eucalyptus area (COS)

(c) Result



This example also produced positive results. Almost all eucalyptus trees were classified (90%), maintaining a relative low missing rate (12%). This area is near the village of Alcobertas, belonging to the municipality of Rio Maior (39.408168, -8.917378).

Hit Rate: 90%
Miss Rate: 12%

Scenario 3 - Isolated and few eucalyptus trees among other land types

In this scenario we discuss two similar examples, one where there are an isolated amount of eucalyptus trees among other types of lands, which depict very well the selectivity of the developed approach. The chosen areas present big diversity of land (roads, eucalyptus trees, non-eucalyptus trees, etc.) and in both cases, the pixels assigned as eucalyptus by the proposed approach are respecting significantly the boundaries of the area delimited by COS (b).

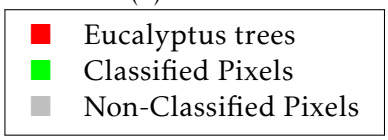
Example 1

This is an area very near the one used in Scenario 1, near the village of Vila Nova de São Pedro, belonging to the municipality of Azambuja (39.195977, -8.799908).

(a) Satellite Image

(b) Eucalyptus area (COS)

(c) Result



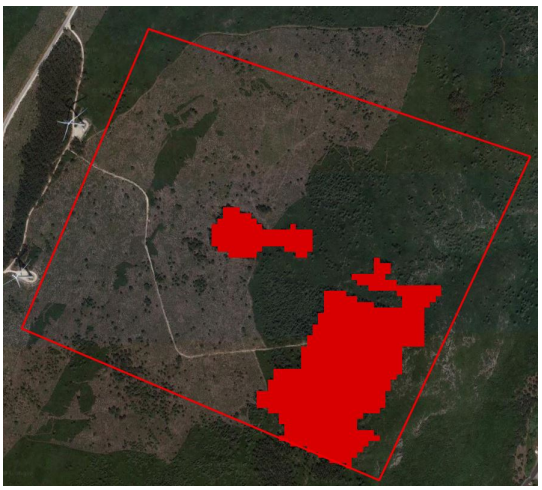
Although it seems our result is not that good, we believe it is the opposite and it is a very interesting result. Observing Figure 5.24 (inside the yellow circle), there is a road that separates two mini eucalyptus forests, but COS marks them as eucalyptus (because it is lower than 1 ha), when they are not. With this result we believe our value-added map will be more accurate for automated classification than COS manual process.

Figure 5.24

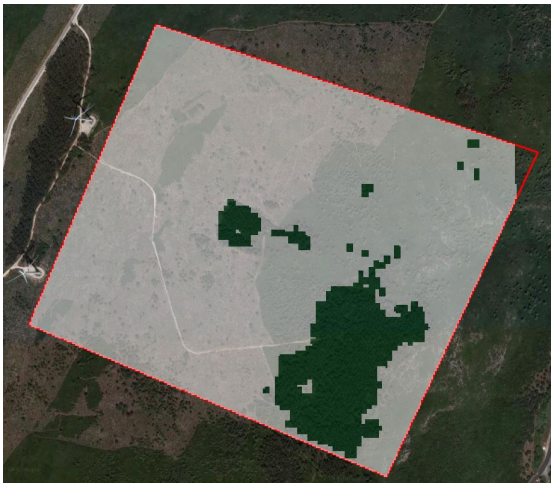
Hit Rate: 71%
Miss rate: 0.01%

Example 2

(a) Satellite Image



(b) Eucalyptus area (COS)



(c) Result

Eucalyptus trees
Classified Pixels
Non-Classified Pixels

This example shares almost the same characteristics as the previous one, i.e. the manual process of COS classified as eucalyptus areas where does not exist eucalyptus . It is from an area in the municipality of Rio Maior (39.385998, -8.938425).

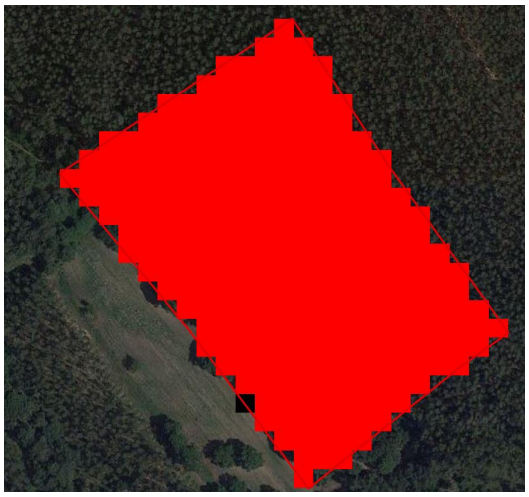
Hit Rate: 76%

Miss Rate: 0.1%

Both examples of scenario (3), although presenting less hit rates when compared with the other ones, again demonstrate that our automated approach can be very useful for the identification of eucalyptus trees, separating elements that are not eucalyptus.

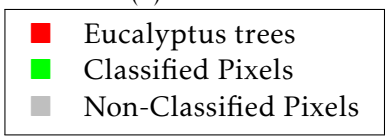
Scenario 4 - Improvement results

(a) Satellite Image



(b) Eucalyptus area (COS)

(c) Result



The last tested area also demonstrates a good improvement in classification of the eucalyptus trees, in relation to COS data. It is an area in the municipality of Ourém (39.694121,-8.586449). In this area, according to COS, there is only one kind of tree (eucalyptus), because they considered anything within 1 ha belongs to the same class.However, by simple visualization we see that gray areas display other types of trees because of 1 ha rule (DGT manual classification). In our case, our automated process already distinguishes eucalyptus from other types of trees, which we believe that it can create a level-3 products (objective of IPSTERS project). Differences between eucalyptus and other trees can be found mainly in their canopy (color) and in their shape. Obviously the hit rate is not very high but we believe our approach is good at distinguishing eucalyptus from

other types of forest and it is an automated process.

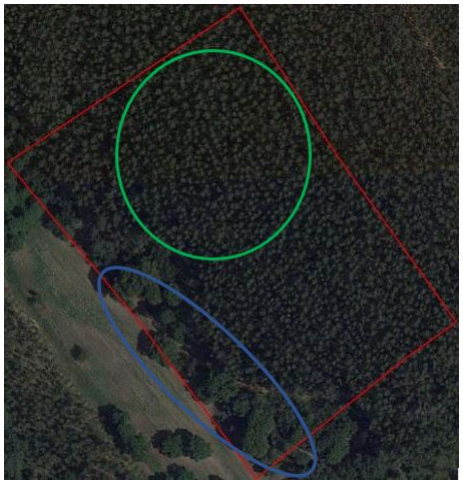


Figure 5.25: Difference between the trees inside blue circle (assumed as not eucalyptus) and green circle (eucalyptus trees).

Hit Rate: 77%

Miss Rate: 0% (like in scenario 1, the miss rate is zero, since all pixels, supposedly, are eucalyptus trees).

Conclusion

As can be seen, the results provided by our approach revealed to be very promising approach regarding land cover classification. Here we tested several scenarios and the results were quite good, with very few misclassified pixels. Eucalyptus areas were proper identified, even when the components of the analysed images are diverse (scenario 2). We believe that this data fusion approach could be a good starting point to further land cover classification and to build level-3 products in value-added maps (IPSTERS project objective).

Chapter 6

Conclusion and Further Work

6.1 Conclusions

In this work, we developed a fuzzy-fusion approach with reinforcement aggregation operators, for land cover classification from multispectral satellite images, more specifically, for eucalyptus trees identification. The main objective of the approach was to fuse spectral information from a multispectral satellite imagery source (Sentinel 2 images) to produce value-added land cover maps, applied to a specific type of forest, the Eucalyptus. As mentioned before, this work was applied to satellite imagery of year 2019 and for the ground truth we got data from COS, provided by DGT.

Several aggregation operators were compared to select which one was the most suitable - for our case study it was the Multiplicative FIMICA, which produced the most consistent and better outputs. As the results and validations suggested, our fuzzy-fusion approach is quite reliable and in some cases surpasses the existent manual method for forestry identification (COS) because it does not classify trees that are not eucalyptus, even though COS classify them as if is (Scenario 4). Further, our automatic data fusion approach is able to separate eucalyptus forests from roads that pass through them, because the ground truth (COS) assumes all area as forest of eucalyptus (scenario 3)

There was also a crucial point regarding the proposed work, the hit/miss rate trade-off. Since the objective of the proposed work is to identify a specific forest type, it can often be confused with another soil type (more specifically other type of vegetation). Thus, careful attention is needed to minimize misclassification, which sometimes influenced the number of hits.

Finally, we also showed how FIF algorithm could be used as a functional mechanism to handle heterogeneous data, normalize it, and produce a fused information aggregation, ready for supporting effective decision making.

Summarizing we believe the data fusion approach, discussed in this work, produced very interesting added-value maps, concerning classification of eucalyptus forest areas, and it seems a robust automated process that can surpass manual classification in some cases.

6.2 Improvements and Future Work

From this study, there are some improvements on the used vegetation indices criteria that could be made:

Poor tree density

In areas where the eucalyptus density is low (i.e. where there is more soil than trees) since we are dealing with multispectral data, sometimes eucalyptus trees are not classified. Improvements on tuning the reinforcement operator and also by adding other criteria (vegetation indices and/or other satellite bands) may solve this problem.

Low tree age

Areas where eucalyptus age is very low, the results and vegetation indices revealed that these areas are treated as soil and are not classified. There is a need to investigate if there are other criteria (vegetation indices) that can be added or created.

As future work we have two aspects. To construct the membership functions with other methods of fuzzification, such as inference, rank ordering and intuition/expert knowledge (Ross 2010), secondly we need to test the g parameter with higher values for the used reinforcement operators (FIMICA and CRO).

As future directions, this work can be easily extended to other kinds of forests or other land types, as a good approach to aid automatic classification/ identification of different land cover.

Bibliography

Ahmed, B. and M. A. Al Noman (2015). "Land cover classification for satellite images based on normalization technique and Artificial Neural Network." In: *2015 International Conference on Computer and Information Engineering (ICCIE)*. IEEE, pp. 138–141.

Ali, I., F. Greifeneder, J. Stamenkovic, M. Neumann, and C. Notarnicola (2015). "Review of machine learning approaches for biomass and soil moisture retrievals from remote sensing data." In: *Remote Sensing* 7.12, pp. 16398–16421.

Ali, S., P Dare, and S. Jones (2008). "Fusion of remotely sensed multispectral imagery and Lidar data for forest structure assessment at the tree level." In: *ISPRS Proceedings, Beijing* 37, B7.

Alves, A. M., J. S. Pereira, and J. M. N. Silva (2007). "A introdução e a expansão do eucalipto em Portugal." In: *Alves, AM, Pereira, JS & Silva, JMN*, pp. 13–60.

Antrop, M. (2004). "Landscape change and the urbanization process in Europe." In: *Landscape and Urban Planning* 67.1. Development of European Landscapes, pp. 9 –26. ISSN: 0169-2046. DOI: https://doi.org/10.1016/S0169-2046(03)00026-4.

Ardeshir Goshtasby, A. and S. Nikolov (2007). "Guest Editorial: Image Fusion: Advances in the State of the Art." In: *Inf. Fusion* 8.2, pp. 114–118. ISSN: 1566-2535. DOI: 10.1016/j.inffus.2006.04.001.

Asner, G. P., J. M. Scurlock, and J. A. Hicke (2003). "Global synthesis of leaf area index observations: implications for ecological and remote sensing studies." In: *Global Ecology and Biogeography* 12.3, pp. 191–205.

Asrar, G., M Fuchs, E. Kanemasu, and J. Hatfield (1984). "Estimating absorbed photosynthetic radiation and leaf area index from spectral reflectance in wheat 1." In: *Agronomy journal* 76.2, pp. 300–306.

Ayhan, E and O Kansu (2012). "Analysis of image classification methods for remote sensing." In: *Experimental Techniques* 36.1, pp. 18–25.

Bannari, A, D Morin, F Bonn, and A. Huete (1995). "A review of vegetation indices." In: *Remote sensing reviews* 13.1-2, pp. 95–120.

Bardallo, J. M., M. A. D. Vega, F. A. Márquez, and A. Peregrín (2015). "Study on the use of uninorm aggregation operators in linguistic fuzzy modeling." In: *IFSA-EUSFLAT*.

Baret, F. (1986). "Contribution au suivi radiométrique de cultures de céréales." Doctoral dissertation. Paris 11.

Bariou, R, D Lecarnus, and P Le Henaff (1985). "Indices de vegetation-Dossiers de télédétéction." In: *Centre Regional de Téléde-téction, Université de Rennes, 150p*.

Baumgardner, M. F., L. F. Silva, L. L. Biehl, and E. R. Stoner (1986). "Reflectance Properties of Soils." In: ed. by N. Brady. Vol. 38. Advances in Agronomy. Academic Press, pp. 1 –44. DOI: https://doi.org/10.1016/S0065-2113(08)60672-0.

Beliakov, G., A. Pradera, T. Calvo, et al. (2007). *Aggregation functions: a guide for practitioners*. Vol. 221. Springer.

Ben-Dor, E., Y. Inbar, and Y. Chen (1997). "The reflectance spectra of organic matter in the visible near-infrared and short wave infrared region (400–2500 nm) during a controlled decomposition process." In: *Remote Sensing of Environment* 61.1, pp. 1 –15. ISSN: 0034-4257. DOI: https://doi.org/10.1016/S0034-4257(96)00120-4.

Benarchid, O. and N. Raissouni (2019). "Potential of Semi-automatic Object-based Land Cover Classifications using Very High Resolution Satellite Images: Tetuan-city Comparison Case Study." In: *Research & Reviews: Journal of Space Science & Technology* 2.3, pp. 19–29.

Bleiholder, J. and F. Naumann (2009). "Data Fusion." In: *ACM Comput. Surv.* 41.1, 1:1–1:41. ISSN: 0360-0300. DOI: 10.1145/1456650.1456651.

Bourdarias, C., P. Da-Cunha, R. Drai, L. F. Simoes, and R. A. Ribeiro (2010). "Optimized and flexible multi-criteria decision making for hazard avoidance." In: *Proceedings of the 33rd Annual AAS Rocky Mountain Guidance and Control Conference*. Citeseer, pp. 5–10.

Calvo, T., G. Mayor, and R. Mesiar (2012). *Aggregation operators: new trends and applications*. Vol. 97. Physica.

Campanella, G. and R. A. Ribeiro (2011). "A framework for dynamic multiple-criteria decision making." In: *Decision Support Systems* 52.1, pp. 52 –60. ISSN: 0167-9236. DOI: https://doi.org/10.1016/j.dss.2011.05.003.

Campbell, J. B. and M. Hall-Beyer (1997). "[Introduction to remote sensing." In: *Cartographica* 34.1, p. 70.

Canziani, G., R. Ferrati, C. Marinelli, and F. Dukatz (2008). "Artificial neural networks and remote sensing in the analysis of the highly variable Pampean shallow lakes." In: *Mathematical Biosciences and Engineering* 5.4, pp. 691–711.

Cerasoli, S., M. Campagnolo, J. Faria, C. Nogueira, and M. d. C. Caldeira (2018). "On estimating the gross primary productivity of Mediterranean grasslands under different fertilization regimes using vegetation indices and hyperspectral reflectance." In: *Biogeosciences* 15.17, pp. 5455–5471.

Chaovalit, P. and L. Zhou (2005). "Movie review mining: A comparison between supervised and unsupervised classification approaches." In: *Proceedings of the 38th annual Hawaii international conference on system sciences*. IEEE, pp. 112c–112c.

Colditz, R. R. (2015). "An Evaluation of Different Training Sample Allocation Schemes for Discrete and Continuous Land Cover Classification Using Decision Tree-Based Algorithms." In: *Remote Sensing* 7.8, pp. 9655–9681. ISSN: 2072-4292. DOI: 10.3390/rs70809655.

Comber, A., P. Fisher, and R. Wadsworth (2005). "What is Land Cover?" In: *Environment and Planning B: Planning and Design* 32.2, pp. 199–209. DOI: 10.1068/b31135.

Coops, N. C., C. Stone, D. S. Culvenor, L. A. Chisholm, and R. N. Merton (2003). "Chlorophyll content in eucalypt vegetation at the leaf and canopy scales as derived from high resolution spectral data." In: *Tree Physiology* 23.1, pp. 23–31. ISSN: 1758-4469. DOI: 10.1093/treephys/23.1.23.

Cyr, L., F. Bonn, and A. Pesant (1995). "Vegetation indices derived from remote sensing for an estimation of soil protection against water erosion." In: *Ecological Modelling* 79.1, pp. 277 –285. ISSN: 0304-3800. DOI: https://doi.org/10.1016/0304-3800(94)00182-H.

Dai, X and S Khorram (1999). "Data fusion using artificial neural networks: a case study on multitemporal change analysis." In: *Computers, Environment and Urban Systems* 23.1, pp. 19 –31. ISSN: 0198-9715. DOI: https://doi.org/10.1016/S0198-9715(98)00051-9.

Dalponte, M., L. Bruzzone, and D. Gianelle (2008). "Fusion of hyperspectral and LIDAR remote sensing data for classification of complex forest areas." In: *IEEE Transactions on Geoscience and Remote Sensing* 46.5, pp. 1416–1427.

DANESHGAR, S. (2015). "Remote sensing observations for monitoring coastal zones: Volturno river mouth case study." In:

Datt, B. (1999). "A New Reflectance Index for Remote Sensing of Chlorophyll Content in Higher Plants: Tests using Eucalyptus Leaves." In: *Journal of Plant Physiology* 154.1, pp. 30 –36. ISSN: 0176-1617. DOI: https://doi.org/10.1016/S0176-1617(99)80314-9.

Davidson, N. and C. Finlayson (2007). "Earth Observation for wetland inventory, assessment and monitoring." In: *Aquatic Conservation: Marine and Freshwater Ecosystems* 17.3, pp. 219–228. DOI: 10.1002/aqc.846.

Dennison, P. E., D. A. Roberts, S. H. Peterson, and J. Rechel (2005). "Use of Normalized Difference Water Index for monitoring live fuel moisture." In: *International Journal of Remote Sensing* 26.5, pp. 1035–1042. DOI: 10.1080/01431160420002739 98.

Desclée, B., P. Bogaert, and P. Defourny (2006). "Forest change detection by statistical object-based method." In: *Remote Sensing of Environment* 102.1, pp. 1 –11. ISSN: 0034-4257. DOI: https://doi.org/10.1016/j.rse.2006.01.013.

Dubois, D. and H. Prade (2012). *Fundamentals of fuzzy sets*. Vol. 7. Springer Science & Business Media.

Eastman, J. R. (2003). "IDRISI Kilimanjaro: guide to GIS and image processing." In:

Emery, W. J., D. Baldwin, and D. Matthews (2003). "Maximum cross correlation automatic satellite image navigation and attitude corrections for open-ocean image navigation."

In: *IEEE Transactions on Geoscience and Remote Sensing* 41.1, pp. 33–42. DOI: 10.1109/TGRS.2002.808061.

Fauvel, M., J. Chanussot, and J. A. Benediktsson (2006). "Decision Fusion for the Classification of Urban Remote Sensing Images." In: *IEEE Transactions on Geoscience and Remote Sensing* 44.10, pp. 2828–2838. DOI: 10.1109/TGRS.2006.876708.

Foody, G. M. (2002). "Status of land cover classification accuracy assessment." In: *Remote sensing of environment* 80.1, pp. 185–201.

Friedl, M. and C. Brodley (1997). "Decision tree classification of land cover from remotely sensed data." In: *Remote Sensing of Environment* 61.3, pp. 399 –409. ISSN: 0034-4257. DOI: https://doi.org/10.1016/S0034-4257(97)00049-7.

Garcia, H. M.V.G. d. et al. (2017). "A floresta em Portugal. Causas e consequências da expansão do Eucalipto. Caso de estudo: O concelho de Torres Vedras." Doctoral dissertation.

Ge, Y., J. A. Thomasson, and R. Sui (2011). "Remote sensing of soil properties in precision agriculture: A review." In: *Frontiers of Earth Science* 5.3, pp. 229–238.

Ghamisi, P., B. Rasti, N. Yokoya, Q. Wang, B. Hofle, L. Bruzzone, F. Bovolo, M. Chi, K. Anders, R. Gloaguen, et al. (2018). "Multisource and Multitemporal Data Fusion in Remote Sensing." In: *arXiv preprint arXiv:1812.08287*.

Gilabert, M., J González-Piqueras, F. Garcıa-Haro, and J Meliá (2002). "A generalized soil-adjusted vegetation index." In: *Remote Sensing of environment* 82.2-3, pp. 303–310.

Gitelson, A. A., Y. J. Kaufman, and M. N. Merzlyak (1996). "Use of a green channel in remote sensing of global vegetation from EOS-MODIS." In: *Remote Sensing of Environment* 58.3, pp. 289 –298. ISSN: 0034-4257. DOI: https://doi.org/10.1016/S0034-4257(96)00072-7.

Goodman, I. R., R. P. Mahler, and H. T. Nguyen (2013). *Mathematics of data fusion*. Vol. 37. Springer Science & Business Media.

Greco, S., R. A. M. Pereira, M. Squillante, and R. R. Yager (2010). *Preferences and decisions: models and applications*. Vol. 257. Springer.

Hao Ying, Yongsheng Ding, Shaokuan Li, and Shihuang Shao (1999). "Comparison of necessary conditions for typical Takagi-Sugeno and Mamdani fuzzy systems as universal approximators." In: *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans* 29.5, pp. 508–514. DOI: 10.1109/3468.784177.

Haywood, A. and C. Stone (2011). "Semi-automating the stand delineation process in mapping natural eucalypt forests." In: *Australian Forestry* 74.1, pp. 13–22. DOI: 10.1080/00049158.2011.10676341.

Hsu, S.-L., P.-W. Gau, I.-L. Wu, and J.-H. Jeng (2009). "Region-based image fusion with artificial neural network." In: *World Academy of Science, Engineering and Technology* 53, pp. 156–159.

HUETE, A. (2004). "11 - REMOTE SENSING FOR ENVIRONMENTAL MONITORING." In: *Environmental Monitoring and Characterization*. Ed. by J. F. Artiola, I. L. Pepper, and

M. L. Brusseau. Burlington: Academic Press, pp. 183 –206. ISBN: 978-0-12-064477-3. DOI: https://doi.org/10.1016/B978-012064477-3/50013-8.

Huete, A. and R. Jackson (1987). "Suitability of spectral indices for evaluating vegetation characteristics on arid rangelands." In: *Remote Sensing of Environment* 23.2, 213 –IN8. ISSN: 0034-4257. DOI: https://doi.org/10.1016/0034-4257(87)90038-1.

Hyder, A. K., E. Shahbazian, and E. Waltz (2012). *Multisensor fusion*. Vol. 70. Springer Science & Business Media.

"Information aggregation in intelligent systems: An application oriented approach" (2013). In: *Knowledge-Based Systems* 38. Special Issue on, pp. 3 –13.

Jamsandekar, S. S. and R. R. Mudholkar (2013). "Self generated fuzzy membership function using ANN clustering technique." In: *International Journal of Latest Trends in Engineering and Technology Special issue IDEAS* 2013, pp. 142–152.

Jassbi, J. J., R. A. Ribeiro, L. M. Camarinha-Matos, J. Barata, and M. I. Gomes (2018). "Continuous Reinforcement Operator applied to Resilience in Disaster Rescue Networks." In: *2018 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pp. 1–7. DOI: 10.1109/FUZZ-IEEE.2018.8491482.

Jin Zhao and B. K. Bose (2002). "Evaluation of membership functions for fuzzy logic controlled induction motor drive." In: *IEEE 2002 28th Annual Conference of the Industrial Electronics Society. IECON 02.* Vol. 1, 229–234 vol.1. DOI: 10.1109/IECON.2002.1187512.

Kanellopoulos, I. and G. G. Wilkinson (1997). "Strategies and best practice for neural network image classification." In: *International Journal of Remote Sensing* 18.4, pp. 711–725. DOI: 10.1080/014311697218719.

Karnieli, A., N. Agam, R. T. Pinker, M. Anderson, M. L. Imhoff, G. G. Gutman, N. Panov, and A. Goldberg (2010). "Use of NDVI and land surface temperature for drought assessment: Merits and limitations." In: *Journal of climate* 23.3, pp. 618–633.

Karnik, N. N., J. M. Mendel, and Q. Liang (1999). "Type-2 fuzzy logic systems." In: *IEEE transactions on Fuzzy Systems* 7.6, pp. 643–658.

Klein, L. A. and L. A. Klein (2004). *Sensor and data fusion: a tool for information assessment and decision making*. Vol. 324. SPIE press Bellingham^ eWA WA.

Klir, G. J. and B. Yuan (1995). "Fuzzy sets and fuzzy logic: theory and applications." In: *Upper Saddle River*, p. 563.

Kosko, B. (1994). "Fuzzy systems as universal approximators." In: *IEEE Transactions on Computers* 43.11, pp. 1329–1333. DOI: 10.1109/12.324566.

Kussul, N., M. Lavreniuk, S. Skakun, and A. Shelestov (2017). "Deep Learning Classification of Land Cover and Crop Types Using Remote Sensing Data." In: *IEEE Geoscience and Remote Sensing Letters* 14.5, pp. 778–782. DOI: 10.1109/LGRS.2017.2681128.

Lavreniuk, M., S. V. Skakun, A. J. Shelestov, B. Y. Yalimov, S. Yanchevskii, D. J. Yaschuk, and A. Kosteckiy (2016). "Large-scale classification of land cover using retrospective satellite data." In: *Cybernetics and Systems Analysis* 52.1, pp. 127–138.

Lee, H., B. Lee, K. Park, and R. Elmasri (2010). "Fusion techniques for reliable information: A survey." In: *International Journal of Digital Content Technology and its Applications* 4.2, pp. 74–88.

Liang, T., X. Sun, H. Wang, R. Ti, and C. Shu (2016). "Airborne polarimetric remote sensing for atmospheric correction." In: *Journal of Sensors* 2016.

Lin, S.-K. (2013). "Introduction to Remote Sensing. By James B. Campbell and Randolph H. Wynne, The GuilfordPress, 2011; 662 pages. Price:£ 80.75, ISBN 978-1-60918-176-5." In: *Remote Sensing* 5.1, pp. 282–283.

Liu, H. and J. Li (2010). "The study of the ecological problems of eucalyptus plantation and sustainable development in Maoming Xiaoliang." In: *Journal of Sustainable development* 3.1, p. 197.

Lunetta, R. S., J. F. Knight, J. Ediriwickrema, J. G. Lyon, and L. D. Worthy (2006). "Land-cover change detection using multi-temporal MODIS NDVI data." In: *Remote Sensing of Environment* 105.2, pp. 142 –154. ISSN: 0034-4257. DOI: https://doi.org/10.1016/j.rse.2006.06.018.

Madry, S. (2013). "Introduction and History of Space Remote Sensing." In: *Handbook of Satellite Applications*, pp. 657–666.

Mamdani, E. and S. Assilian (1975). "An experiment in linguistic synthesis with a fuzzy logic controller." In: *International Journal of Man-Machine Studies* 7.1, pp. 1 –13. ISSN: 0020-7373. DOI: https://doi.org/10.1016/S0020-7373(75)80002-2.

Manyika, J. and H. Durrant-Whyte (1995). *Data Fusion and Sensor Management: a decentralized information-theoretic approach*. Prentice Hall PTR.

Mas, J. F. and J. J. Flores (2008). "The application of artificial neural networks to the analysis of remotely sensed data." In: *International Journal of Remote Sensing* 29.3, pp. 617–663. DOI: 10.1080/01431160701352154.

Matsushita, B., W. Yang, J. Chen, Y. Onda, and G. Qiu (2007). "Sensitivity of the enhanced vegetation index (EVI) and normalized difference vegetation index (NDVI) to topographic effects: a case study in high-density cypress forest." In: *Sensors* 7.11, pp. 2636–2651.

McFEETERS, S. K. (1996). "The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features." In: *International Journal of Remote Sensing* 17.7, pp. 1425–1432. DOI: 10.1080/01431169608948714.

McNairn, H. and R. Protz (1993). "Mapping Corn Residue Cover on Agricultural Fields in Oxford County, Ontario, Using Thematic Mapper." In: *Canadian Journal of Remote Sensing* 19.2, pp. 152–159. DOI: 10.1080/07038992.1993.10874543.

Mohamad, A (2016). "A New Image Contrast Enhancement in Fuzzy Property Domain Plane for a True Color Images." In: *International Journal of Signal Processing Systems* 4.1, pp. 45–50.

Monteiro Alves, A., J. Pereira, and J. Silva (2007). "A introdução e a expansão do eucalipto em Portugal." In: pp. 13 –24. ISBN: 978-972-8669-25-6.

Moodley, A. (2016a). "Language Identification With Decision Trees: Identification Of Individual Words In The South African Languages." Doctoral dissertation. Bachelor's Thesis, University of South Africa.

— (2016b). "Language Identification With Decision Trees: Identification Of Individual Words In The South African Languages." Doctoral dissertation. DOI: 10.13140/RG.2.2.25539.81445.

Mora, A., J. M. Fonseca, and R. Ribeiro (2013). "Real-time image recovery using temporal image fusion." In: *2013 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pp. 1–5. DOI: 10.1109/FUZZ-IEEE.2013.6622539.

Mora, A. D., A. J. Falcão, L. Miranda, R. A. Ribeiro, and J. M. Fonseca (2015). "A fuzzy multicriteria approach for data fusion." In: *Multisensor Data Fusion From Algorithms and Architectural Design to Applications*, pp. 109–126.

Mora, A., T. M. A. Santos, S. Łukasik, J. M. N. Silva, A. J. Falcão, J. M. Fonseca, and R. A. Ribeiro (2017). "Land Cover Classification from Multispectral Data Using Computational Intelligence Tools: A Comparative Study." In: *Information* 8.4. ISSN: 2078-2489. DOI: 10.3390/info8040147.

Mukherjee, I. and S. Routroy (2012). "Comparing the performance of neural networks developed by using Levenberg–Marquardt and Quasi-Newton with the gradient descent algorithm for modelling a multiple response grinding process." In: *Expert Systems with Applications* 39.3, pp. 2397–2407.

Myint, S. W., P. Gober, A. Brazel, S. Grossman-Clarke, and Q. Weng (2011). "Per-pixel vs. object-based classification of urban land cover extraction using high spatial resolution imagery." In: *Remote sensing of environment* 115.5, pp. 1145–1161.

Nedeljkovic, I. (2004). "Image classification based on fuzzy logic." In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 34.30, pp. 3–7.

O'Brien, M. A. and J. M. Irvine (2004). *Information fusion for feature extraction and the development of geospatial information*. Tech. rep. NATIONAL GEOSPATIAL-INTELLIGENCE AGENCY (NGA) RESTON VA.

Ouchi, K. (2013). "Recent trend and advance of synthetic aperture radar with selected topics." In: *Remote Sensing* 5.2, pp. 716–807.

Pais, T. C., R. A. Ribeiro, and L. F. Simões (2010). "Uncertainty in dynamically changing input data." In: *Computational Intelligence in Complex Decision Systems*. Springer, pp. 47–66.

Pal, M. and P. M. Mather (2003). "An assessment of the effectiveness of decision tree methods for land cover classification." In: *Remote sensing of environment* 86.4, pp. 554–565.

Perry, C. R. and L. F. Lautenschlager (1984). "Functional equivalence of spectral vegetation indices." In: *Remote Sensing of Environment* 14.1, pp. 169 –182. ISSN: 0034-4257. DOI: https://doi.org/10.1016/0034-4257(84)90013-0.

Phillips, S. (2002). "Reducing the computation time of the Isodata and K-means unsupervised classification algorithms." In: *IEEE International Geoscience and Remote Sensing Symposium*. Vol. 3. IEEE, pp. 1627–1629.

Piella, G. (2003). "A general framework for multiresolution image fusion: from pixels to regions." In: *Information Fusion* 4.4, pp. 259 –280. ISSN: 1566-2535. DOI: https://doi.org/10.1016/S1566-2535(03)00046-0.

Piiroinen, R., J. Heiskanen, E. Maeda, A. Viinikka, and P. Pellikka (2017). "Classification of Tree Species in a Diverse African Agroforestry Landscape Using Imaging Spectroscopy and Laser Scanning." In: *Remote Sensing* 9.9. ISSN: 2072-4292. DOI: 10.3390/rs9090875.

Pisani, R. J., R. Y. M. Nakamura, P. S. Riedel, C. R. L. Zimback, A. X. Falcao, and J. P. Papa (2014). "Toward satellite-based land cover classification through optimum-path forest." In: *IEEE Transactions on Geoscience and Remote Sensing* 52.10, pp. 6075–6085.

Pohl, C. and J. L. Van Genderen (1998). "Review article multisensor image fusion in remote sensing: concepts, methods and applications." In: *International journal of remote sensing* 19.5, pp. 823–854.

Qi, J, M. MORAN, A. HUETE, R. Jackson, and A Chehbouni (1991). "View-atmosphere-soil effects on vegetation indices derived from SPOT images." In: *ESA, Physical Measurements and Signatures in Remote Sensing*, 2.

Qiao, H., M. Wu, M. Shakir, L. Wang, J. Kang, and Z. Niu (2016). "Classification of small-scale eucalyptus plantations based on NDVI time series obtained from multiple high-resolution datasets." In: *Remote sensing* 8.2, p. 117.

Ribeiro, R. A., T. C. Pais, and L. F. Simões (2010). "Benefits of full-reinforcement operators for spacecraft target landing." In: *Preferences and Decisions*. Springer, pp. 353–367.

Ribeiro, R. A., A. Falcão, A. Mora, and J. M. Fonseca (2014). "FIF: A fuzzy information fusion algorithm based on multi-criteria decision making." In: *Knowledge-Based Systems* 58. Intelligent Decision Support Making Tools and Techniques: IDSMT, pp. 23 –32. ISSN: 0950-7051. DOI: https://doi.org/10.1016/j.knosys.2013.08.032.

Ribeiro, R. A. (1996). "Fuzzy multiple attribute decision making: A review and new preference elicitation techniques." In: *Fuzzy Sets and Systems* 78.2. Fuzzy Multiple Criteria Decision Making, pp. 155 –181. ISSN: 0165-0114. DOI: https://doi.org/10.1016/0165-0114(95)00166-2.

Ribeiro, R. A. and R. A. M. Pereira (2003). "Generalized mixture operators using weighting functions: A comparative study with WA and OWA." In: *European Journal of Operational Research* 145.2, pp. 329 –342. ISSN: 0377-2217. DOI: https://doi.org/10.1016/S0377-2217(02)00538-6.

Richards, J. A. and J. Richards (1999). *Remote sensing digital image analysis*. Vol. 3. Springer.

Rocchini, D., C. Ricotta, and A. Chiarucci (2007). "Using satellite imagery to assess plant species richness: The role of multispectral systems." In: *Applied Vegetation Science* 10.3, pp. 325–331. DOI: 10.1111/j.1654-109X.2007.tb00431.x.

Rouse Jr, J. W., R. H. Haas, J. Schell, and D. Deering (1973). "Monitoring the vernal advancement and retrogradation (green wave effect) of natural vegetation." In:

Rudas, I., I. Batyrshin, A. Zavala, O. Camacho, and L Villa Vargas (2009). "Digital fuzzy parametric conjunctions for hardware implementation of fuzzy systems." In: pp. 157 –166. DOI: 10.1109/ICCCYB.2009.5393942.

Rudas, I. J., E. Pap, and J. Fodor (2013). "Information aggregation in intelligent systems: An application oriented approach." In: *Knowledge-Based Systems* 38, pp. 3–13.

Santos, T. M. A., A. Mora, R. A. Ribeiro, and J. M. N. Silva (2016). "Fuzzy-fusion approach for land cover classification." In: *2016 IEEE 20th Jubilee International Conference on Intelligent Engineering Systems (INES)*, pp. 177–182. DOI: 10.1109/INES.2016.7555116.

Schmitt, M. and X. X. Zhu (2016). "Data fusion and remote sensing: An ever-growing relationship." In: *IEEE Geoscience and Remote Sensing Magazine* 4.4, pp. 6–23.

Simonetti, E., D. Simonetti, and D. Preatoni (2014). "Phenology-based land cover classification using Landsat 8 time series." In: *European Commission Joint Research Center: Ispra, Italy*.

Soliman, H., A.-F. Attia, M Hellal, and M. Badr (2006). "Power System Stabilizer Driven by an Adaptive Fuzzy Set for Better Dynamic Performance." In: *Acta Polytechnica* 46.2.

Sowmya, D, P Deepa, and K Venugopal (2017). "Remote Sensing Satellite Image Processing Techniques for Image Classification: A Comprehensive Survey." In: *International Journal of Computer Applications* 161.11, pp. 24–37.

Stoklasa, J., J. Talasová, and P. Holeček (2011). "Academic Staff Performance Evaluation – Variants of Models." In: *Acta Polytechnica Hungarica* 8, pp. 91–111.

Sumfleth, K. and R. Duttmann (2008). "Prediction of soil property distribution in paddy soil landscapes using terrain data and satellite information as indicators." In: *Ecological Indicators* 8.5, pp. 485 –501. ISSN: 1470-160X. DOI: https://doi.org/10.1016/j.ecolind.2007.05.005.

Taylor, J. C., T. R. Brewer, and A. C. Bird (2000). "Monitoring landscape change in the National Parks of England and Wales using aerial photo interpretation and GIS." In: *International Journal of Remote Sensing* 21.13-14, pp. 2737–2752. DOI: 10.1080/01431160050110269.

Theiler, J. P. and G. Gisler (1997). "Contiguity-enhanced k-means clustering algorithm for unsupervised multispectral image segmentation." In: *Algorithms, Devices, and Systems for Optical Information Processing*. Vol. 3159. International Society for Optics and Photonics, pp. 108–118.

Timothy, J et al. (2010). *Fuzzy logic with engineering applications*.

Torra, V. and Y. Narukawa (2007). "Information fusion and aggregation operators." In: *Cognitive Technologies. Springer, Heidelberg*.

Toth, C. and G. Jóźków (2016). "Remote sensing platforms and sensors: A survey." In: *ISPRS Journal of Photogrammetry and Remote Sensing* 115, pp. 22–36.

Toth, C. and G. Jóźków (2016). "Remote sensing platforms and sensors: A survey." In: *ISPRS Journal of Photogrammetry and Remote Sensing* 115. Theme issue 'State-of-the-art in photogrammetry, remote sensing and spatial information science', pp. 22 –36. ISSN: 0924-2716. DOI: https://doi.org/10.1016/j.isprsjprs.2015.10.004.

Triantaphyllou, E. (2000). "Multi-criteria decision making methods." In: *Multi-criteria decision making methods: A comparative study*. Springer, pp. 5–21.

TTakagi, M. (1985). "Fuzzyidentificationofsystemsanditsapplicationstomodelingandcontrol." In: *IEEETransactionson System, Man, Cybernetics* 15.1, 116r132.

Tzeng, G.-H. and J.-J. Huang (2011). *Multiple attribute decision making: methods and applications*. Chapman and Hall/CRC.

Waltz, E., J. Llinas, et al. (1990). *Multisensor data fusion*. Vol. 685. Artech house Boston.

Weber, S. (1983). "A general concept of fuzzy connectives, negations and implications based on t-norms and t-conorms." In: *Fuzzy Sets and Systems* 11.1, pp. 115 –134. ISSN: 0165-0114. DOI: https://doi.org/10.1016/S0165-0114(83)80073-6.

Weih, R. C. and N. D. Riggan (2010). "Object-based classification vs. pixel-based classification: comparative importance of multi-resolution imagery." In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 38.4, p. C7.

Wilson, R. (2013). *Advanced remote sensing: terrestrial information extraction and applications, by Shunlin Liang, Xiaowen Li and Jindi Wang: Oxford, Academic Press, 2012, 799 pp.,£ 95 (hardback), ISBN 978-0-12-385954-9*.

Xu, H. (2006). "Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery." In: *International journal of remote sensing* 27.14, pp. 3025–3033.

Xu, M., P. Watanachaturaporn, P. K. Varshney, and M. K. Arora (2005). "Decision tree regression for soft classification of remote sensing data." In: *Remote Sensing of Environment* 97.3, pp. 322 –336. ISSN: 0034-4257. DOI: https://doi.org/10.1016/j.rse.2005.05.008.

Yager, R. R. and A. Rybalov (1998). "Full reinforcement operators in aggregation techniques." In: *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 28.6, pp. 757–769. ISSN: 1083-4419. DOI: 10.1109/3477.735386.

Yager, R. R. (1980). "On a general class of fuzzy connectives." In: *Fuzzy Sets and Systems* 4.3, pp. 235 –242. ISSN: 0165-0114. DOI: https://doi.org/10.1016/0165-0114(80)90013-5.

Yan, G., J.-F. Mas, B. Maathuis, Z. Xiangmin, and P. Van Dijk (2006). "Comparison of pixel-based and object-oriented image classification approaches—a case study in a coal fire area, Wuda, Inner Mongolia, China." In: *International Journal of Remote Sensing* 27.18, pp. 4039–4055.

Yang, S., Q. Feng, T. Liang, B. Liu, W. Zhang, and H. Xie (2018). "Modeling grassland above-ground biomass based on artificial neural network and remote sensing in the

Three-River Headwaters Region." In: *Remote Sensing of Environment* 204, pp. 448 –455. ISSN: 0034-4257. DOI: https://doi.org/10.1016/j.rse.2017.10.011.

Yu, Q., P. Gong, N. Clinton, G. Biging, M. Kelly, and D. Schirokauer (2006). "Object-based detailed vegetation classification with airborne high spatial resolution remote sensing imagery." In: *Photogrammetric Engineering & Remote Sensing* 72.7, pp. 799–811.

Zadeh, L. A. (1988). "Fuzzy logic." In: *Computer* 21.4, pp. 83–93. DOI: 10.1109/2.53.

Zadeh, L. (1965). "Fuzzy sets." In: *Information and Control* 8.3, pp. 338 –353. ISSN: 0019-9958. DOI: https://doi.org/10.1016/S0019-9958(65)90241-X.

Zadeh, L. A. (1995). "Discussion: Probability Theory and Fuzzy Logic Are Complementary Rather Than Competitive." In: *Technometrics* 37.3, pp. 271–276. DOI: 10.1080/00401706.1995.10484330.

Zimmermann, H.-J. (2011). *Fuzzy set theory—and its applications*. Springer Science & Business Media.